

QY	481	ggaggaaaaagaaataagcaaaacaaccccgctttctcggactcggccttattcccat	540
Db	481	GGAGGAAAAAGAAATAGCAAAAACAACCCCGTTTCTGGACTGGGCGCTTATTCACATTTAT	540
QY	541	tggagcaatctccgcttgcgcggagagatgaaagaanaatactcgtttgtgtaaaaagacac	600
Db	541	TGGACCAATGTCGGTTGGCGGAGATGAAGAAAATATCTGCTTGTGTAAAAAGACATC	600
QY	601	tggcagggttggggtgtgtcctcccaagaagatcgcctgtctcaacgltatgctttcccat	660
Db	601	TGGCAGGGTGGGGTGTGTCTCCACAAGATGCGACACTGCTCTCACTGTAGCTTTTCCAT	660
QY	661	ggcccaacgctccccaatcattcgcctctgtcgtgaggcaagcagtgctcaataagccgggtg	720
Db	661	GGCCCAACGCTCCCAATCATTCATTCGCGCTGTGGAGGCGACAGTGTGCATGAAGCCGGG	720
QY	721	gagcctcctcaatgctctggccagttggggagaccggtttgtatgaccaatgggatatacgcat	780
Db	721	GAGCCTCAATGCTCTGGCCAGTGGGGAGCCGTTGTATGATCAACAAAGGATATGCCGAT	780
QY	781	gcaagagatgacitgcaggcagctggcgctcagttgacitgcacaaacatcggcatcaagga	840
Db	781	GCAGAGATGATCTCGACGGACACTGGCGCTAGTGGCATTTGCCAAACATGCGATCAGCA	840
QY	841	tattttgggaaagggtcgtgcacagltatgtgatagaagtaagtcgtgcactcgtggaatgag	900
Db	841	TATTTTGGGAAAGGGTCTGGCGCCAGTTATGTTGGATGAAGTACGCTGCATGGGAATGAG	900
QY	901	cttccaattggcaggtgtccaaagaagtcctcggggagagataacgtctgcataagaa	960
Db	901	CTTCAATTTGACAGTGTCCAAAGAGCTCTGGGGAGAGCATATACGTGGCCATTAAGAA	960
QY	961	gatcgtgagatgctccitgtaccccctcaacagatcgggtctcaagctcagttgggaa	1020
Db	961	GATCGTGAGTGTCTCTGTACCCCTCTACAGATGGGGTCTACGACTTCAGTGGGAAA	1020
QY	1021	ggcagccaatgaaagggtcgccttggagataltatcacagagccagctggygaaatcgtgtgat	1080
Db	1021	GGCAGCCATGAGGGGTGCGCTTGGAGAGTATTTACAGAGGCCAGTGGGAGACTGTCTGTAT	1080
QY	1081	gatggcttggactcgtcgtgaataacatacgtgtgttgcgaacagttgggatttaataatgct	1140
Db	1081	GATGGCTGGACTGTGAGTAATACATACGTGTGTTCGACAGTGGGATTTAAATATGCT	1140
QY	1141	aaacaagcattctgcacaacattttgaaagaagaacagggccatacgtgttgaataagct	1200
Db	1141	AAACAAGCATTGGCCAAACATTTTGAAGAAAGCACAGGGCCCATATGTGTGATATACGTC	1200
QY	1201	agctgtcctcaagaaagaaacacagaattctctcagtgctcgaagcgacagtggygaaagcat	1260
Db	1201	AGCTGTCTCAGGAAAGGAACACAGATTCTTCACATGTTCCAGGGCACAGTGGGAAAGCAT	1260
QY	1261	gactgcagccaccccggaagaatgttaagatactgctcgtctaccccttggcgcgaaagacaacag	1320
Db	1261	GACTGCAGCCACCCCGGAAGATGTTAGATTGACCTGCTACCTCGCGCCAGGAGAACAGG	1320
QY	1321	ctctcctcggggttttccctgtcagaactatgtatcgtgaaagaaataagaaagaaagcaggtg	1380
Db	1321	CTCTCTCTGGGTTTCTCTGTCAAGCTATGTGATGGGAAAAATTAAGAAAGAACAGCAGTG	1380
QY	1381	gaggttttatacfaatgycagctggggaacacatctgtatgtatgtatgaltgagctatagaagt	1440
Db	1381	GAGGTTTATATCAATGCGCCAGTGGGGAACAATCTGTGATGATGAGTGAATGAAGAT	1440
QY	1441	gcagcgtgtgacitcgtcgtcagcttggctacaagaggtcctcgcagagcaagaacatgct	1500
Db	1441	GCAGCTGTGATCTGTCTCTCAAGCTTGGCTTACAAAGGCTCTCCAGAGCAAGAACATGAGCT	1500
QY	1501	tactttggagaagaaagaaagcaatccatcctgtgtataatgtgaagtgtgcacgggaatgag	1560
Db	1501	TACTTTGGAGAAGAAAGCAACCATTCATGTTGATAATGTAAAGTGCACAGGAATATAG	1560

QY	1561	aggtcttggtgagctgctatcaagcaaatgattggagaacaaactgcgcgcacagctgaa	1620
Db	1561	AGGTCCTTGGGTGAGCTGATCAACCAAGATTGGAAAGACAAACTGCCGCCACAGTGA	1620
QY	1621	gatgcagaagctatcttgattatcttggcaagaagggctcaagctaaacgtataaagag	1680
Db	1621	GATGAGAGAGTTATTGTGATTATTTTGGCAGAGAGGCTCTCAGTAAACGTAATTAAG	1680
QY	1681	tcctctcatcctgttcttggtccttgaattactgcacgctcggcagaagcggatcatgtt	1740
Db	1681	TCCCTCATCTGATTGTGGGTTGAGATTACTGCACGTCGGCAGAGCGGATCATGGT	1740
QY	1741	gggaaaaatcttcttaaggggtgtgttgcccttggcaggttccctcggctgaatcatcc	1800
Db	1741	GGGAAAAATTTCTTAAGGGGGGTGGCTTGGCAGAGTTTCCCTCCGGCTGAAGTCAAC	1800
QY	1801	catgtagatgtagcagcgcctctcgagggcctcctcctcctagtagctgtgtgctctaca	1860
Db	1801	CATGAGATGGCAGAGCTCTCTTGGGGGCTACGCTCTAGTAGAGTGTGCTGGCTCTCA	1860
QY	1861	gcagacacactgtttcaagaggatgtagcaacagagactagagctatgctgttaaggttga	1920
Db	1861	GCAGACACACTGTTCAAGAGGATAGGAACAGACACTAGAGCTATGCTTAGGGTTGA	1920
QY	1921	gattatcatctctgtgtacagagagatttgagaagaaattggagcttcaacagattgtg	1980
Db	1921	GATTATCATCTACTCTGATACCAAGAGAGATTGAGGAATATGGAGTTCAACACATGTG	1980
QY	1981	attatctggaggtatgtagcccgacccgacgtatgattatgatatagacccttgtagattaca	2040
Db	1981	ATTCTATCGGAGTATGCAACCCGACCGCAGTATATGATATAGATAGCCCTGGTTAGATTCA	2040
QY	2041	ggaaccagaagagcaatgtgcacagattcagogaacatgttttcacagcctgtttacaactc	2100
Db	2041	GGACACAGAAGCAATGTGCCAATTCAGAGAGCAATGTTTGCCAGCCTGTTTACACTC	2100
QY	2101	tggagagagagccacagaaacagcatccaaactgttatacaacagatgggtgtgacaca	2160
Db	2101	TGGAGAGAGAGGCCACAGAAACAGCATCTCAACTGTATCATATACAGATGGGGTGACACA	2160
QY	2161	ggagagagcctatcaagaacactaacaagaagcaccatcccttacttccctaaagttc	2220
Db	2161	GGAGAGAGCTATTCTAAGAACACTCAACAACAGACGCCATTCCTTACTCTTAAGAGTTT	2220
QY	2221	tgtgaagaagcttataagagtggtgttcaagagagaagatgctgtgtcctgggaacctccat	2280
Db	2221	TGTGAAGAGCTTATAAGGGTTCGGTTTACAGSAGAGATCCTTTGTGTCTGGAAACTCCAT	2280
QY	2281	gaacacaaacgctgtgaacagctgcagaggagacagcggagacacatcattgtgaaocg	2340
Db	2281	GAACACAAACCGGTGAGACACTGCCAGGGGAGACAGGGGAGGCCACTCATGTGTGAACGG	2340
QY	2341	cccgagagagagcttgggtgtgtgtatgggtgtagctcctctgggggtataggtctgtggagtc	2400
Db	2341	CCCGGAGAGAGCTGGGTGTGTGTGGGGGTAGCTCCTCGGGGTATGGCTGTGGAGTCAAG	2400
QY	2401	gattccctgtgttcttaccaaagctcaagccttgtatcctctgtataaagtgtaacc	2460
Db	2401	GATTCTCCTGCTGTTTATACCAAGCTCAGCCCTTGTATCCTTGGATTAATAAAGTGTACC	2460
QY	2461	aaactgtaatctctcatgtagaacttcaaaagcagcatlttaacaaatggaaaacttgaac	2520
Db	2461	AAACTGTATTTCTTCAAGAACTTCAAGCAGCATTTTAACAAATGTAATTTTGAAC	2520
QY	2521	ccccactattagcaactcagcagagatgtagacaacaaacggcgaag	2562
Db	2521	CCCCACTATTAGCACTCAGCAGAGATGACAAACAAACGGCAGAAG	2562

QY 841 taatttggggaaggtctgcccagttatgttgatgaagtaagctgcaatgggaatgag 900
 Db 1043 TATTtggggaaggtctgcccagttatgttgatgaagtaagctgcaatgggaatgag 1102
 QY 901 ctctcaattgagagtgctcaagaagctcctggggaagagcaactgttgccataaagaa 960
 Db 1103 CTTCATTGAGCAAGTGTCCAAAGAGCTCCTGGGAGAGCAATAGTGCCCATTAAGAA 1182
 QY 961 gatctggaatgtctctgtaccctctaacagatgggttcatcagactgtcagtgaggaa 1020
 Db 1163 GATGCTGGAATGCTCTTACCCCTTAACAGATGGGGTCATCAGACTTGCAGGTGGGAA 1222
 QY 1021 ggcagccatgaggtctgcttggaagtataatacagagccagtgaggaaatgtctgtat 1080
 Db 1223 GGCAGCCATAGAGGTCTCTTGGAGGTATATACAGAGCCAGTGGGGAACGTGTGTAT 1282
 QY 1081 gatgctgagctgagctggaatatacagtggtgtgtgagagtggttaataatggt 1140
 Db 1283 GATGCTGAGCTGAGTGAATACATACGTGTGTGTGACAGTGGGATTAATATAGT 1342
 QY 1141 aaacaagcatctgcacaacatttgaaagaagcagagcccatatgttgatgagctc 1200
 Db 1343 AACCAAGCATCTGCCAACATTTTGAAGAAAGCAGAGGCCATATGTTGATGAGCTC 1402
 QY 1201 agctgtcaggaagaagaacacagatctctcaagtgttccagagcagtgaggagcat 1260
 Db 1403 AGCTGCTCAGGAAGAGAACAGATTCTCAGTGTCCAGGCAAGTGGGGAAGCAT 1462
 QY 1261 gactgagccagccggaagaatgttagatgtcgtcagctcagccgagcagagcagag 1320
 Db 1463 GACTGCAACCCAGCCAGAAATGTTAGCATTTGCTTACCTGGGCGAGAGGACAGG 1522
 QY 1321 ctctctctgggttctcctgtcagactgtatgataagaaataaagaagaagcagtg 1380
 Db 1523 CTCTCTCTGGGTTTCTCTGTACAGACTGATGATGAGAAATAAGAAAGAGCAGTG 1582
 QY 1381 gagggtttatcaatggccagtgagggaacaactgtgtatgataagtgactgtataagat 1440
 Db 1583 GAGGTTTATCAATAGCCAGTGGGGAACAATCTGTATGTGATGAGACTGATTAAGAT 1642
 QY 1441 gcaagctgtatctgtcagctgtgactgtacaaaggtctcagagcaagaacatgct 1500
 Db 1643 GCAGCTGATCTGTCTGCTACGCTTGGCTACAAAGGCTCTCCAGAGCAAGACATGGCT 1702
 QY 1501 tacttggagaagaagaagaccatccatcagtgatgataatgtgaagtgcacaggaatgag 1560
 Db 1703 TACTTGGAGAAAGAAAGAACCATTCATGTGATGATGAAAGTGACAGAAATGAG 1762
 QY 1561 aggtcttggctgactgtatcaagaagataatgtgaagaacaactgtccgccaagtga 1620
 Db 1763 AGGTCTTGGCTGACTGTATCAAGCAAGATTTGGAAGACCAACTCCCGCACAGTGA 1822
 QY 1621 gatcgaggaattatgtgatatatttggcaagaagccctcaggttaacagtaataaagag 1680
 Db 1823 GATGCAAGAGATTATTTGTGATTTTGGCAAGAGCCCTCAGTAACAGTAATAAGAG 1882
 QY 1681 tccctctcatctgttgtgactgtgagataactgacagctcagcagaagcagatcatggt 1740
 Db 1883 TCCCTCTCATCTGTTGTGTGCTTGAATTAAGTACCGTGGGCAAGAGCGATCATTTGCT 1942
 QY 1741 ggggaaatctttaaaggggtgtgtgcttggcaggtttccctcggctgtgaatcatcc 1800
 Db 1943 GGGAAATCTTTAAGGGGTGTGTGGCTTGGAGAGTTTCCCTCCGGCTAAATCATATCC 2002
 QY 1801 catgaggaatgagagctcctcctgagcagagctcagagctcagtaactgtgtgtctaca 1860
 Db 2003 CATGGAGATGAGAGCTCTCTGGGGCTACGCTCTGATACCTCTGGTCTTCA 2062
 QY 1861 gcaagcactgttctcaagaaggtatgcaacagcactagagactgtgttaaggttga 1920
 Db 2063 GCAGCAGACTGTTCAGAGAGTATGCAACAGACTAGAGCTATGCTTTAGGGTTGA 2122
 QY 1921 gattatcatctctgttaccagagaggttggaggaagaatgtgagttcaacagattgtg 1980

Db 2123 GATTATCATCTGTGATACAGAGAGATTGAGAGAAATTGAGACTTAACAGATTGTG 2182
 QY 1981 attcatcggaatcatcaccagcagccagtgatgatgataagcctgtgttatgataa 2040
 Db 2183 ATTCTATGAGATGATGACCCGAGCCAGTGTATGACATAGCCTGGTTGATTAACA 2242
 QY 2041 ggaacagaagaagaatgtgccaagattcagcagccatgttttgcagcctgtttacac 2100
 Db 2243 GACACAGAAAGCAATGTGCAGATTACAGAGCATGTTTGGCAGCTGTATACACATC 2302
 QY 2101 tggagagaagagccagcaagaagaacagcatcctaactgtttacataacagatgggtgacaa 2160
 Db 2303 TGGAGAGAGAGCCACAGAAACAGATCCAACTGTATTAACAGAGAGGGGTGACACA 2362
 QY 2161 ggaagagccattcaagaacatacaacaagagccatcccttacttacttaaaagttt 2220
 Db 2363 GAGCAGACTTATTCAGACACTACACAAAGCAGCCATTCCTTACTTCTTAAGGTTT 2422
 QY 2221 tgtgaagaagttataaggttgcgtttacagaggagaatgttctgttctgaaactccat 2280
 Db 2423 TGTGAAGACGTTATTAAGGTGCGTTTACAGGAGATGCTTGTCTGGAACCTCCAT 2482
 QY 2281 gaacacaacgcgttgacagctcagcaggaagacagcggagaccatcatgtgtgaag 2340
 Db 2483 GAACACAAACGCGTGGACACTCCAGGAGACAGCGAGACACTCATGTGTGAACG 2542
 QY 2341 cccgagagagctgtgtgtgtatggtgtgactcctcctgggggtatgtgtgagtcag 2400
 Db 2543 CCGGAGAGAGCTGGGTGGTGTATGGGGTACCTCTGGGGGTATGGCTGTGAGTCAAG 2602
 QY 2401 gatctcctgtgtttataacaaagtcagccttctgacttgaataaaggtgtacc 2460
 Db 2603 GATTCTCTGTGTATATACAAAGTCTCAGCTTGTACTGTGATTAAGATGTCACC 2662
 QY 2461 aaactgaaattcttcataagaaacttcaagaagcagcatttaacaaatgaattgaac 2520
 Db 2663 AAACGTAAATCTTCAAGAACTTCAAGACACATTTAAACAAATGAAGAACTTGAAC 2722
 QY 2521 ccccatattagcactcagcagagatgacacaaagcaag 2562
 Db 2723 CCCCATTTAGCACTCAGAGATGACACAAAGGCAAG 2764

RESULT 3
 X19027 standard: DNA: 2614 BP.
 AC X19027;
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3 encoding DNA.
 KW Serine protease; BSSP-3; brain tissue; ss.
 OS Mus sp.
 PN M05905290-AT
 PD 04-FEB-1999.
 PE 24-JUL-1998: J03324.
 PR 24-JUL-1997: JF-213969.
 RA (SDNR)-SUNTORX-LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 DR P-PSDB; W99088.
 PT New serine protease expressed in brain tissue - used in screening
 for potential serine protease inhibitors for drug use
 PS Example 1; Page 45-50; 69pp; Japanese.
 CC The present sequence encodes a serine protease designated BSSP-3, which
 is isolated from mouse brain tissue. Transformants may be used to
 produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SO Sequence 2614 BP; 614 A; 671 C; 799 G; 530 T;

Query Match 54.6%; Score 1398; DB 1; Length 2614;
 Best Local Similarity 84.2%; Pred. No. 0;

FT CDS 24..2309
 FT /tag- b
 FT sig_peptide 24..86
 FT /tag- c
 FT mat_peptide 87..2306
 FT /tag- d
 FT misc_feature 90..275
 FT /tag- e
 FT /note- "proline rich, basic segment"
 FT misc_feature 276..494
 FT /tag- f
 FT /note- "Kringles domain"
 FT misc_feature 519..824
 FT /tag- g
 FT /note- "SRCR domain 1"
 FT misc_feature 840..1142
 FT /tag- h
 FT /note- "SRCR domain 2"
 FT misc_feature 1179..1484
 FT /tag- i
 FT /note- "SRCR domain 3"
 FT misc_feature 1536..2306
 FT /tag- j
 FT /note- "proteolytic domain"
 FT misc_feature 1707..1709
 FT /tag- k
 FT /note- "histidine of the catalytic triade"
 FT misc_feature 1857..1859
 FT /tag- l
 FT /note- "aspartic acid of the catalytic triade"
 FT misc_feature 2154..2156
 FT /tag- m
 FT /note- "serine of the catalytic triade"
 FT 3'UTR 2307..2356
 FT /tag- n
 FT polyA_signal 2324..2329
 FT /tag- o
 FT polyA_signal 2331..2336
 FT /tag- p
 FT polyA_signal 2357..2376
 FT /tag- q
 FT MO9849322-A1.
 FT 05-NOV-1998
 FT 24-APR-1998..180625..
 FT 26-APR-1997: CH-000966
 FT (SOMD)/ SONDREGER P.
 FT SONDREGER P.
 FT WPI: 99-009438/01.
 FT P-PSDB: W83362.
 FT New human and murine neurotrophin - used, e.g. for inhibiting
 FT neurite growth, treatment of neurological or lung disease, including by
 FT gene therapy and in drug development.
 FT Claim 1: Page 29-32: 50pp: English.
 FT The present sequence encodes mouse neurotrophin. Neurotrophin proteins
 FT and polynucleotides can be used: (i) to inhibit tumours, including
 FT metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 FT caused by stroke or brain injury (having a protective effect on the
 FT penumbra zone); (iii) to treat or prevent neurodegeneration, e.g.
 FT neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 FT increase survival of damaged neurons (e.g. in cases of hypoxia,
 FT ischaemia, nerve transection) and to stimulate regeneration and/or
 FT restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 FT degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 FT causes of cell death) in the nervous system; (vii) to regenerate brain
 FT and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 FT performance, including learning and memory; (x) to treat or prevent a
 FT wide range of psychiatric disorders; and (xi) to treat brain or lung
 FT injury associated with protease expression (specifically emphysema or
 FT bronchitis).
 FT Sequence 2376 BP; 589 A; 586 C; 708 G; 493 T;

Query Match 54.4%; Score 1393.2; DB 1; Length 2376;
 Best Local Similarity 84.5%; Pred. No. 0;
 Matches 1566; Conservative 0; Mismatches 288; Indels 0; Gaps 0;
 QY 618 tccctcaagaatgagcagctgctgacgctgcttccctccctgagcccaagttcccat 677
 Db 458 TGCCCAAGGCGAAGTAGACTGGGCTACTGCTCCATTGGTTCAGAGCCCGGCTGCCGT 517
 QY 678 cattgccttctgtgagcagcagctgtgcatgaagccgggtgagctcaccatgctg 737
 Db 518 CATTGGCCTTGTGTGGGGAACAGTGGGATGATGAGTGCAGTGGAGCTTACCAAGCTGG 577
 QY 738 ccagttggggaacccgtttgcatgacccaatggagtgatgctccgatcagaagtatccag 797
 Db 578 CCAAGTGGGGACCATCTGTGCACCAATGGGCAATGGGCAATGGGCAATGGGCAATGGG 637
 QY 798 gcaagctggcctcagctgagccttcccaagcagtcagcctcagccttctggggaagggct 857
 Db 638 GCAAGCTGGGCTCAGTGGCTTCCCAAGCATGCTCAGGCAATTTGGGGAAGGATC 697
 QY 858 tggcccaagttaigtgtgataagtaacgctgacatgaggaatgagcttcaattgagcagtg 917
 Db 698 TGGCCCAATATGTGTGAGTAAGTACGTGCAACCGAAGAGCTGTCAATTTGAGCAATG 757
 QY 918 tccaagagctcctggtggagagcataactgtggtccataaagaagatgctgaggttctgt 977
 Db 758 TCCAAGAGTCTCTGGGGGGAATCACTGTGCCATTAAGAGATGCTGAGAGTCTTGTG 817
 QY 978 taacccctcaagatgagtggtcctcagcctcagctgagctgaggaagcagcagagagctg 1037
 Db 818 TGTTCCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 877
 QY 1038 ctggaagatataaagaagccagtggtgggaactgctgtgatagtgctgagcagcagct 1097
 Db 878 CTTGAGAGTCTACTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 937
 QY 1098 gaatacaatgctgt 1157
 Db 938 GAACACATAGT 997
 QY 1158 ccaatttgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1217
 Db 998 CCAATTTGATGAGCAGACAGAGGCGCATATGCTGTGATGCTGTGATGCTGTGATGCT 1057
 QY 1218 aaccagattcttcaatgcttccagcagcagcagcagcagcagcagcagcagcagcagc 1277
 Db 1058 AGTCAGCTTATTCATGATGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
 QY 1278 agattttagcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1337
 Db 1118 AGATGTGGGGCTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1177
 QY 1338 tgcacacagatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1397
 Db 1178 CATCAGACTAGTGTGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
 QY 1398 ccagttggggaacaatctgtgagatgagatgagatgagatgagatgagatgagatgag 1457
 Db 1238 CCAATGGGGAACAATCTGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
 QY 1458 tcaagcttgctacaagaggtctgccaagagcagaacacatgcttctcttggagaagaa 1517
 Db 1298 GCAACTTGGCTATTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357
 QY 1518 agagccatccatgctgagatgagatgagatgagatgagatgagatgagatgagatgag 1577
 Db 1358 AGGCCCATTCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1417
 QY 1578 tatcaagcaagatattgagagacacacacacacacacacacacacacacacacacacac 1637
 Db 1418 TGTCAACAAGACATGTGAGAGGACACACTGCGCCACAGTGAAGATGAGAGATGATG 1477
 QY 1638 tgattatttggcagaagagcctcagtgtaacagtaataaagaagctccctcattctgtt 1697

Db 1238 GTCCATATGGCTGATGAGTATTTTGTTCCTCCCAAGAGTCATCCATTGAAGAGTCA 1297
QY 920 caaagagctcctctgggagagcataactgtgccaataaagaagatgctgagtgctctgta 979
Db 1298 GAATTAGACAGTGGGTGTGTGAGAGCTGTTCGCACGACGAAGATGCTGGGGTCACTTGCA 1357
QY 980 c 980
Db 1358 c 1358

Search completed: February 27, 2000, 03:42:40
Job time: 2697 sec


```

GENERAL INFORMATION:
APPLICANT: Li, Xiao
APPLICANT: Snyder, Solomon H
TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
TITLE OF INVENTION: Protein Associated with Taste Buds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,350B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mollfe, Susan A
REGISTRATION NUMBER: 33,568
REFERENCE/DOCKET NUMBER: 01107,48790
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4360 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: 94..3963
US-08-470-350B-1

```

```

Query Match          3.8%; Score 97.8; DB 1; Length 4360;
Best Local Similarity 57.3%; Pred. No. 4.5e-17;
Matches 177; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
1330 gggttcctcgtcagctgatgatgagaaataagaagaagcagtgaggtttt 1389
1330 gggttcctcgtcagctgatgatgagaaataagaagaagcagtgaggtttt 1389
958 GGTTCGCTGTGAGCGTGATGAGAGACAGTGTCAGGCTGCTGAGACATCCTT 1017
1390 atcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1449
1018 TACCAAGGCTTCCTGGGGTACGCTGTGACGACAGCTGGGACACCAAGAG 1077
1450 atctgcctcagctgagcagcagcagcagcagcagcagcagcagcagcagc 1509
1078 GCTGTCAGGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
1510 gaagaaagaaagcagcagcagcagcagcagcagcagcagcagcagcagc 1569
1138 CAAGCTCTGATCACTGCTGATGATGATGATGATGATGATGATGATGATG 1197
1570 gctgactgctcagcagcagcagcagcagcagcagcagcagcagcagcag 1629
1198 TGGAGCTGCTCCACCGAGCTGCTCTCTCACTCACTGATGATGATGATG 1257
1630 gttattgt 1638
1258 GTCATCTGT 1266

```

```

RESULT 3
US-08-477-674-9
Sequence 9, Application US/08477674
Patent No. 5644035
GENERAL INFORMATION:
APPLICANT: Kohls, Kirston E.
APPLICANT: Halebeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,674
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE:
APPLICATION NUMBER: US/07/961,404
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2285 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-674-9

```

```

Query Match          3.7%; Score 94.4; DB 1; Length 2285;
Best Local Similarity 56.4%; Pred. No. 2.7e-16;
Matches 176; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
678 cattcgcttcgtcagcagcagcagcagcagcagcagcagcagcagcagc 737
248 CATCGGCTGCGCGATGCGGCGCCACCAACAGGCGCGCGGATTTTACAGAG 307
738 ccagctgggaacccgttcgtgataccaatggatgatgatacagcagcagc 797
308 CCACTGGGCGACTGTGTGACCAACTGTGTGACCTGATGATGATGATGATG 367
798 gcaagctggcctcagctgagcagcagcagcagcagcagcagcagcagcag 857
368 GGCCTGGGCTTGAGAAAGCCACCGAGCTGCGGACAGAGCTGCTGCGGCA 427
858 tggccagctatgttgagaaagcagcagcagcagcagcagcagcagcagcag 917
428 AGGCCCATATGCTGTGAGAGGTCCAGTGCACGGGAAACCGAGCGCTCTG 487

```


STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 47..1402
US-08-453-117-1

Query Match 3.7%; Score 93.8; DB 1; Length 2028;
Best local similarity 56.2%; Pred. No. 3.7e-16;
Matches 176; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 668 cgttccccatcattcgctctgtgagagcagctgtgcatgaagccggttgaagctct 727
1083 CATTTACGAAGTTGCGACTGTGCTGGGAGCGCCCTCACGAGGGGAGGTGAGATAC 1142
728 accatgctggccagtgaggaaaccgtttgtgatgaaccaatgggagtgtccgatgcagaag 787
DB 1143 TCCACAGCGGCCCGTGGGGTACAAATTGTGACGATGCTGGGAAGTGCCTGTTGACAGG 1202
QY 788 tgatctgcagggcagctgggcccagtgagcatgccaagcatgcatcagcatatttg 847
DB 1203 TCCTCTGTAGGAGCTTGCGATACCGAGGTTCAGCCGTGCACAAAGCAGCTCACTTTG 1262
QY 848 gggaaagggtctggcccagttatgttgaatgaagtaacgtgcactgggaaatgaagcttcaa 907
DB 1263 GACAAAGTACTGTGCTCAATATGCGCATGAACTGTTTGTGTTGGAGAGATCATCTA 1322
QY 908 ttgagcagtgctcaagaagctcctctggggagagcataactgtggccataaagaagatgctg 967
DB 1323 TTGAAGAAATGTAAATTCGCAATGGGGACAAAGACCTGTTTCACATTCCTGAAGATGCTG 1382
QY 968 gaatgctctgtac 980
DB 1383 GAGTCACTTGAC 1395

Search completed: February 27, 2000, 21:33:35
Job time: 177 sec

C	8	329.2	12.8	632	60	A1808920	wf66h04.x
C	9	317	12.4	440	42	A1152507	u92a204.x
C	10	314.2	12.3	513	31	AA313027	EST183845
C	11	314	12.3	483	44	A1325429	ms46a12.x
C	12	297.4	11.6	426	49	A1653707	wp36d08.x
C	13	296.4	11.6	448	63	A1970451	wr10h10.x
C	14	295.2	11.5	449	49	A1624687	ts43f05.x
C	15	271	10.6	274	32	AA373034	EST84994
C	16	233.4	9.1	384	28	AA063841	ml81f05.x
C	17	223.6	9.1	379	62	A1893676	ml98c11.y
C	18	224.6	8.8	365	28	AA073513	ml98c11.y
C	19	223.2	8.7	303	43	A1195404	ue72h11.x
C	20	206.8	8.1	354	44	A1327025	ml98c11.x
C	21	181.2	7.1	347	102	A0343432	PC111-12
C	22	152	5.9	402	33	AA417058	zui3c05.x
C	23	127.8	5.0	181	20	T07472	EST05361.Fe
C	24	125.4	4.9	230	70	AV252179	AV252179
C	25	123	4.8	268	50	AV038271	AV038271
C	26	118.8	4.6	163	27	AA023924	mb92d01.x
C	27	117.2	4.6	423	37	AA063308	ae90d07.s
C	28	116.6	4.6	196	70	AV228999	AV228999
C	29	116.4	4.5	424	74	AW206795	UI-H-B11-
C	30	111.6	4.4	215	70	AV247069	AV247069
C	31	106.6	4.2	466	26	W68576	z34c12.xl
C	32	101.8	4.0	599	74	AA184423	fj14c03.y
C	33	99.2	3.9	599	74	A1652526	w61d11.x
C	34	98.8	3.9	705	81	B16308	34707.TP.CI
C	35	98.2	3.8	570	79	GA7140105	GA7140105
C	36	96.4	3.8	545	26	W88611	zh73g06.xl
C	37	95.4	3.7	460	49	A1047474	DXFZ5860
C	38	95	3.7	411	46	A1449832	mt74d08.x
C	39	93.8	3.7	641	51	A1734941	at05d08.x
C	40	92.8	3.6	474	35	AA581208	nd38d06.x
C	41	92.2	3.6	455	50	A1677867	wd34d04.x
C	42	90.6	3.5	382	28	Z81180	SSCF12.POT
C	43	89.6	3.5	525	82	AO676495	HS-5512.A
C	44	88.6	3.5	616	45	A1341167	q89n023.x
C	45	86.2	3.4	468	46	A1415453	mc57d06.x

ALIGNMENTS

RESULT 1
A1968581/C
LOCUS
DEFINITION
A1968581 684 bp mRNA
wt90c08.x1 NCI-CGAP GC6 Homo sapiens CDNA clone IMAGE:2514734 3'
similar to TR:008762 008762 PROTEASE, SERINE, 12 NEUROTRIPSIN, ;
mRNA sequence.

ACCESSION
A1968581
A1968581.1 GI:5765399
EST.
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 684)

REFERENCE
AUTHORS
TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
On Jun 22, 1998 this sequence version replaced gi:3246664.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

JOURNAL
COMMENT
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone-distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:

		www.bio.lim.gov/db/rrp/image/image.html	
		Seq primer: -40UP from g1bco	
		High quality sequence stop: 410.	
FEATURES		Location/Qualifiers	
source		1..684	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2514734"	
		/clone_lib="NCI-CGAP GC6"	
		/tissue_type="pooled germ cell tumors"	
		/lab_host="PH10B"	
		/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 147592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		164 a 175 c 141 g 204 t	
ORIGIN			
Query Match		23.5%; Score 602.6; DB 63; Length 684;	
Best Local Similarity		98.3%; Pred. No. 4.2e-134;	
Matches 630; Conservative		0; Mismatches 9; Indels 2; Gaps 2;	
QY	1924	tatcatctctgtgtaccagagaggttggaggaatgtgagattgaacatttgatt	1983
DB	684	TACCATTTTGGGTACCCAGAGAGATTGAGGAAGAAATTCACAGATGGGAT	625
QY	1984	catcgagagatcgaccgcgcgc-gcagtgattatgacatagccctgttagattcaag	2042
DB	624	CATCGGAGATCGAACCGACCGCAGCTATTGACATAGCCTTGTTAGATTCAAG	565
QY	2043	accagaagcaatgtgcagatttcagagcca-tgttttcagagcctgtttaccattc	2101
DB	564	ACCAGAAGCAATGTGCCAGATTTCAGAGCCCAATGTTTCCAGCGCTTTACCACT	505
QY	2102	ggagagagagccacaggaacagatccactgttaccataacagatgggtgacacag	2161
DB	504	GGAGAGAGAGCCACAGAAACAGCATCCACTGTATACAGAGATGGGTGACACG	445
QY	2162	gacgagccttcaagaacactcaacacagcagccattcccttactcctaaagtttt	2221
DB	444	GACGAGCCTTTCAAGACACTCAACACAGACCCATTCCCTTACTCTTAAGGTTT	385
QY	2222	gtgaagaagctataaaggtcggtttaccagggagaatgtctgtgtgaaacctccat	2281
DB	384	GTCAGAGACCTATTAAAGGTCGGTTTACAGGGAGATGCTTGCTGTAACCTCATG	325
QY	2282	aacacaaacgctgtgacagcttcgagggagacagcggaagaccattcatgttgaagc	2341
DB	324	AACACAAACCGTGACAGCTGCCAGGGAGACACCGGAGACCACTCATGTGACGCG	265
QY	2342	ccggaagaagctgt	2401
DB	264	CCGGAAGAGACTGGGT	205
QY	2402	atttcctgtgtttataccaaagcttcagccctgttacctgtgtataaaagtgcacca	2461
DB	204	ATTTCCTGTGTATTATACCAAGCTCTCAGCCTTGTACTTGGAATAAAAGTGACACA	145
QY	2462	aactgttaattcttcaggaactcaaacagcatttaacaaatggaacttgaacc	2521
DB	144	AACGTGAATCTTCATGAGAACTTCAAGACAGATTAAACAAATGAAACTTGAAC	85
QY	2522	cccactatgagcactcagcagagatgacacaaacagcagcag	2562
DB	84	CCCACATTTGACACTCAGCAGAGATGACACAAATGGCAG	44

RESULT 2
LOCUS AI457628/c
DEFINITION tJ53e07.x1 Soares_NSF_F8_9W_OT_PA.P_S1 Homo sapiens cDNA clone IMAGE:2145212 3' similar to TM:008762 008762 NEUROTRPFPSIN PRECURSOR ; , mRNA sequence.
ACCESSION AI457628
VERSION AI457628.1 GI:4310497
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 614)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgp.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3138391.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LINTL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1113 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 384.

FEATURES
Source
Location/Qualifiers
1..614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2145212"
/clone.lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneids: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 1450382-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HR-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739980-740999 Subtraction by Benicio Soares and M. Fatima Bernaldo."

BASE COUNT 146 a 138 c 124 g 204 t 2 others

ORIGIN

Query Match 16.4%; Score 419.8; DB 46; Length 614;
Best Local Similarity 95.8%; Pred. No. 2.6e-90;

Matches 430; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 2114 cacagaaaacagcatcacaactgttaccatcaagaattggtgtgacacagcagctatt 2173
| | | | |
Db 613 CCCAGAAAACAGCATCCANCCTGTTCAATACACGATGGGTGCACACAGGCCAGCCTATT 554
| | | | |
OY 2114 caaagaacctcaaaacacgaacccattcccttaactcccataaaggtttgtgaagaactt 2233
| | | | |
Db 553 CAAAGAACCCTCAACAAACAGCATTCCTTCCTTAATCTCNTPAAAGGTTTGTAAGAACGTT 494
| | | | |
OY 2234 ataaagtgcggtttcacaggagaatgctttgtcgtgaaacctccatatcaacaacgcg 2293
| | | | |
Db 493 ATAAGGGTCGGTTCACAGGAAGAAATGCTTTGTGCTGGAACCTCATGAACAACAACGG 434
| | | | |
OY 2294 ttgacagctgccaggagagacagcgagagacccaatcatgtgtgaacgcccggagagagct 2353
| | | | |

Db	433	TGACAGCTGCAGGAGAGACACGGAGCACACTCATGTGTGAACGGCCCGAAAAAACT	374
Qy	2354	gggttggtatgtaggttgcctcctctgggggatagcgctgtggaaactaaggattcttcctgtg	2413
Db	373	GGGTGTGTATGGGTGACCTCCTCTGGGGTATAGCGTGAGATCAAGGATTTCCTCGGGT	314
Qy	2414	ttataccaagaatcccaagccttgcaccttgatataaaaagtgcaccaacacgttaattc	2473
Db	313	TTTATACCAAATCTCAGCGCTTGTGCCCTTGATATAAAGGTTCCCCAAAGCTATTTT	254
Qy	2474	tcatggaacctccaagcagcatttaacaatgtaaaacattgaacccccactttgc	2533
Db	253	TCATGGAACCTCCAAGCAGCATTTAAACAATGGAATAAATTGAACCCCTTATTAGC	194
Qy	2534	actgcagagatgacacaacaacggcaag	2562
Db	193	ACTCAGCAAGAATGCCAACAAATGGCAAG	165
RESULT	3		
LOCUS	AAA60641	397 bp	mRNA EST 09-JUN-1997
DEFINITION	zfx2d12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:796055 5' similar to SW-TRF2_CANFA P06872 TRYPSINOGEN,		
ACCESSION	AAA60641		
VERSION	AAA60641.1	GI:2185761	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 397) Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kuuba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,K. and Wilson,R. WashU-Merck Est Project 1997 Unpublished (1997)		
JOURNAL	On Sep 12, 1996 this sequence version replaced gi:1394273.		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. Seq primer: -28ml3 rev2 EM from Amershams High quality sequence stop: 273.		
FEATURES	Location/Qualifiers		
SOURCE	1..397		
	/organism="Homo sapiens"		
	/db_xref="GDB:6039851"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:796055"		
	/clone_lib="Soares_total_fetus_Nb2HF8_9w"		
	/dev_stage="8-9 weeks"		
	/lab_host="DH10B"		
	/note="Vector: pUT3D-Pac (Pharmacia) with a modified polynker. Site.1: Not I; Site.2: Eco RI; 1st strand was prepared from mRNA obtained from pooled 8-9 week fetal testis material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pUT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	115 a	87 c	103 g 92 t
ORIGIN			

/clone.lib="Soares_NFL.T.GBC.S1"
/lab.host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19w, testis NHT, and B-cell NCI-CGAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

140 a 130 c 106 g 161 t

Query Match 13.2%; Score 337; DB 69; Length 537;

Best Local Similarity 93.7%; Pred. No. 1.7e-70; Mismatches 20; Indels 6; Gaps 3;

Matches 385; Conservative 0;

QY 2152 ggtgacacggagagccattcaagaacactcaacagcagccttcctactct 2211
DB 537 GGGGACACAGGACGACCATTT--AAGAACCTAAACAGCACATTCCTTACT-CT 483
QY 2212 aaaaagtttctgtaagaacgttataaggtcgtgtttacaagggaaatgtgtctgga 2271
DB 482 TAAAGGTTTGAAGACGTTA-AAGGTCGGTTTACAGGGAATGCTTGTGTGGA 424
QY 2272 aacctcattgaacacaaacgctgtgacagctgcacaggagacagcgagaccactcaty 2331
DB 423 AACCTCATGAAACAAACCCGTGGACAGCTGCCAGGAGACGCGAGACCATCATG 364
QY 2332 tctgacacggccggagagagctgtgtgtatggtgagagccctgggggtatgtctgt 2391
DB 363 tctgacacggccggagagagctgtgtgtatggtgagagccctgggggtatgtctgt 304
QY 2392 ggaagtcagaagatctctgtgtttataccaaagctcagcctttgacttgatgaataa 2451
DB 303 GGAGTCAAGGATTCTCTGTTTATACCAAGTCTCAGCCTTGTACTTGATGAATAA 244
QY 2452 agtgtaccacaaactgaattcttcattggaacttcaagcagcatttaacaaatgaa 2511
DB 243 AGGTGACCAAACTGTAATCTTCATGGAACCTTCAACAGCATTTAAACAATGGAAA 184
2512 actttgaaccccatattagcactcagcagagatgacacaaacggcgaag 2562
183 ACTTTGAACCCCATATTAGCAGCTCAGCAGATGACAAACAATGGCAAG 133

RESULT 6

LOCUS AA928660 478 bp mRNA EST 07-JUL-1998

DEFINITION OM75908.s1 NCI_CGAP_GC4 Homo sapiens CDNA clone IMAGE:1533054 3'

sequence. similar to TR:008762 008762 NEURORTRYPSPIN PRECURSOR. ; mRNA

ACCESSION AA928660
VERSION AA928660.1 GI:3076951

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 478)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2045775.

COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_L_Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the 1.M.A.G.E. Consortium/BLNL at:

www.bio.lnlnl.gov/bbrp/image/image.html

Insert Length: 616 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 382.
Location/Qualifiers

FEATURES

SOURCE

1..478

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1553054"

/clone.lib="NCI_CGAP_GC4"

/tissue_type="pooled germ cell tumors"

/lab.host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from 3 pooled

germ cell tumors, and was then primed with a Not I

oligo(dt) primer. Double-stranded cDNA was ligated to Eco

RI adaptors (Pharmacia), digested with Not I and cloned

into the Not I and Eco RI sites of the modified pT73

vector. Library is normalized. Library was constructed by

Bento Soares and M. Fatima Bonaldo."

BASE COUNT

131 a 123 c 90 g 134 t

ORIGIN

Query Match 13.0%; Score 333.8; DB 40; Length 478;

Best Local Similarity 99.1%; Pred. No. 9.7e-70; Mismatches 2; Indels 1; Gaps 1;

Matches 346; Conservative 0;

QY 2215 aggttttgaagaacgttataaggtcgtgtttacaaggaagatgttgcgtgaac 2274
DB 478 AGGTTTGTGAAGAACGTTATTAAGGTCGTTTACAGGAGAACTGCTGTGCTGGAAC 419
QY 2275 ctccatgaacacaaacgctgtgaca-gctgcagaggagacagcgagacactatgtg 2333
DB 418 CTCCTATGAACAAACAGTGTGACAGCTGCCAGGAGACGCGAGACCATCATGTG 359
QY 2334 tgaacggcccgagagagctgtgtgtatggtgagctcctctgggggtatggtctgtg 2393
DB 358 TGAACGGCCCGAGAGAGCTGTGCTGTATGGGTGACTCTCTGGGGTATGGCTGTGG 299
QY 2394 agtcaaggaattctctgtgtttataccaagctcagccttctgtacttgataaaaag 2453
DB 298 AGTCAAGGATTCCTCGTGTATTATACCAAGTCTCAGCCTTGTACCTTGATTAAG 239
QY 2454 tctcaccacaaactgaattcttcattggaacttcaagcagcatttaacaaatgaaac 2513
DB 238 TCTCACCAAAAGTGAATTTTCATGGAACCTTCAAGCGCATTTAAACAATGGAAAAC 179
QY 2514 ttgtgaaccccatattagcactcagcagagatgacacaaacggcgaag 2562
DB 178 TTTGAACCCCATATTAGCAGCTCAGCAGATGACAAACAATGGCAAG 130

RESULT 7

LOCUS AA166524 470 bp mRNA EST 19-DEC-1996

DEFINITION ms46a12.r1 life Tech mouse embryo 13 5dpc 10666014 Mus musculus

CDNA clone IMAGE:614590 5' similar to SW:UROT_MOUSE P11214 TISSUE

PLASMINOGEN ACTIVATOR PRECURSOR ; mRNA sequence.

ACCESSION AA166524
VERSION AA166524.1 GI:1744123

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

WIDEOR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-p protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:09:43 2000; Maspar time 18.99 Seconds
560.722 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (1-822) from US09147947A.pep (1 of 2)
Perfect Score: 6328
Sequence: 1 PPTRRPPPPRRPPRRPALP.....PGVTKVSAFVPMIKSVTKL 822

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1
Statistics: Mean 35.408; Variance 155.210; scale 0.228

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length	ID	Description	Pred. No.
1	603	9.5	638	2	US-08-681- Sequence 3, Applicatio	5.78e-44
2	597	9.4	356	2	US-08-681- Sequence 1, Applicati	2.12e-43
3	583	9.2	355	2	US-08-811- Sequence 47, Applicati	4.37e-42
4	580	9.2	430	1	US-07-942- Sequence 3, Applicati	8.35e-42
5	583	9.2	437	2	US-08-811- Sequence 51, Applicati	4.37e-42
6	583	9.2	527	2	US-08-811- Sequence 39, Applicati	4.37e-42
7	573	9.1	355	1	US-08-427- Sequence 6, Applicatio	3.79e-41
8	573	9.1	355	4	5232356-1 Patent No. 5233256.	3.79e-41
9	573	9.1	355	1	US-08-427- Sequence 2, Applicatio	3.79e-41
10	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
11	573	9.1	355	1	US-08-137- Sequence 1, Applicatio	3.79e-41
12	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
13	573	9.1	355	1	US-08-217- Sequence 1, Applicatio	3.79e-41
14	573	9.1	355	2	US-08-811- Sequence 45, Applicati	3.79e-41
15	573	9.1	356	1	US-08-427- Sequence 8, Applicatio	1.04e-41
16	573	9.1	356	1	US-08-427- Sequence 4, Applicatio	3.79e-41
17	578	9.1	365	1	US-08-720- Sequence 83, Applicati	1.29e-41
18	578	9.1	363	1	US-08-093- Sequence 83, Applicati	1.29e-41
19	578	9.1	393	2	US-08-560- Sequence 44, Applicati	1.29e-41
20	578	9.1	411	1	US-08-153- Sequence 18, Applicati	1.29e-41
21	578	9.1	411	1	US-08-087- Sequence 1, Applicatio	1.29e-41
22	578	9.1	411	1	US-08-286- Sequence 18, Applicati	1.29e-41
23	578	9.1	431	4	5188829-1 Patent No. 5188829.	1.29e-41

RESULT	1	STANDARD:	PRT:	638 AA.
ID	US-08-681-151-3			
XX	xxxxxx			
XX				
DT				
XX				
DE	Sequence 3, Application US/08681151			
CC	Sequence 3, Application US/08681151			
CC	Patent No. 5869637			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Au-Young, Janice			
CC	APPLICANT: Bandman, Olga			
CC	APPLICANT: Braxton, Scott Michael			
CC	APPLICANT: Goll, Surya			
CC	TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN			
CC	NUMBER OF SEQUENCES: 4			
CC	CORRESPONDENCE ADDRESSES:			
CC	ADDRESSEE: INCYTE PHARMACEUTICALS, INC.			
CC	STREET: 3174 Porter Drive			
CC	CITY: Palo Alto			
CC	STATE: CA			
CC	COUNTRY: US			
CC	ZIP: 94304			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Diskette			
CC	COMPUTER: IBM Compatible			
CC	OPERATING SYSTEM: DOS			
CC	SOFTWARE: FastSeq Version 1.5			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/681,151			
CC	FILING DATE: Herewith			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER:			
CC	FILING DATE:			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Billings, Lucy J.			
CC	REGISTRATION NUMBER: 36,749			
CC	REFERENCE/DOCKET NUMBER: PF-0074US			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: 415-855-0555			
CC	TELEFAX: 415-845-4166			
CC	TELEX:			
CC	INFORMATION FOR SEQ ID NO: 3:			
CC	SEQUENCE CHARACTERISTICS:			

CC Sequence 51, Application US/08811949
CC Patent No. 5840533
CC GENERAL INFORMATION:
CC APPLICANT: NIWA, MINEO
CC APPLICANT: SAITO, YOSHIMASA
CC APPLICANT: SASAKI, HITOSHI
CC APPLICANT: HAYASHI, MASAKO
CC APPLICANT: NOTANI, JOUJI
CC APPLICANT: KOBAYASHI, MASAKAZU
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
CC NUMBER OF SEQUENCES: 67
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 437 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE: 437 AA; 48707 MW; 965246 CN;
SQ
Query Match 9.2%; Score 583; DB 2; Length 437;
Best Local Similarity 37.6%; Pred. No. 4.37e-42;
Matches 114; Conservative 66; Mismatches 101; Indels 22; Gaps 18;
Db 136 LGKHNCRNPDGDAKPMCHVKNRRLTWECVPS-CSTGGLROYSQPOFRITGGLFADI 134
531 IGRHN-CRHS-DAGVICYF-GKKAS-GNSNKESLSSVCGRLRLHRRQRIIGKNSLR 586
195 ASHPQQAIFAKHRSPGGRFLCGGILISSCWITLSAACHFOERF--PPHLLTVILGRYYR 252
587 GGPMPQVSLRLKSSHG-D-RLCGATLLSSCWLTAAHCK-RYGNSTRSYAVRG-DYH 643
Db 253 -VYPGEEOKEFEVKEYIVKREDDDTYDNDIALLOKSDSSRCAQSSSVTVCLP-AD 310
644 TLVPEEFEEIGVQVIVHREYRPDRSDYDIALVRLQGPEDCARFSSHVLPACLPMLRE 703
Db 311 LQLPWTECELSGYKHEALSPFYSERLKEAHVRLYPSSRCTSOHLNRTVTDNMLCAGD 370
704 RPKQTASNCYITIGWG--DT-GRAYSTRLOQAALPLPKRFC-ERYKGR-FTGRMLCAGN 758
Db 371 TRSGGQANLHADCGDSGGLVCLNDGAMTLV-GIISWGLCGGOKDVPGYTKVTNYLD 429
759 LHE--HKRV-DSCGDSGGLPMLCERPGESWVYGYTSMYGGCGVADSPGYTKVSAFVP 814
Db 430 WIR 432
815 WIR 817
RESULT 6

ID US-08-811-949-39 STANDARD; PRT; 527 AA.
XX
XX xxxxxx
XX
AC
DT
XX
XX
DE Sequence 39, Application US/08811949
CC Sequence 39, Application US/08811949
CC Patent No. 5840533
CC GENERAL INFORMATION:
CC APPLICANT: NIWA, MINEO
CC APPLICANT: SAITO, YOSHIMASA
CC APPLICANT: SASAKI, HITOSHI
CC APPLICANT: HAYASHI, MASAKO
CC APPLICANT: NOTANI, JOUJI
CC APPLICANT: KOBAYASHI, MASAKAZU
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
CC NUMBER OF SEQUENCES: 67
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC CITY: ARLINGTON
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22202
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/811,949
CC FILING DATE: 05-MAR-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OBLON, NORMAN F.
CC REGISTRATION NUMBER: 24,618
CC REFERENCE/DOCKET NUMBER: 18-966-0
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-413-3000
CC TELEFAX: 703-413-2220
CC INFORMATION FOR SEQ ID NO: 39:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 527 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE: 527 AA; 59027 MW; 1397140 CN;
SQ
Query Match 9.2%; Score 583; DB 2; Length 527;
Best Local Similarity 37.6%; Pred. No. 4.37e-42;
Matches 114; Conservative 66; Mismatches 101; Indels 22; Gaps 18;
Db 226 LGKHNCRNPDGDAKPMCHVKNRRLTWECVPS-CSTGGLROYSQPOFRITGGLFADI 284
531 IGRHN-CRHS-DAGVICYF-GKKAS-GNSNKESLSSVCGRLRLHRRQRIIGKNSLR 586
Db 285 ASHPQQAIFAKHRSPGGRFLCGGILISSCWITLSAACHFOERF--PPHLLTVILGRYYR 342
587 GGPMPQVSLRLKSSHG-D-RLCGATLLSSCWLTAAHCK-RYGNSTRSYAVRG-DYH 643
Db 343 -VYPGEEOKEFEVKEYIVKREDDDTYDNDIALLOKSDSSRCAQSSSVTVCLP-AD 400
644 TLVPEEFEEIGVQVIVHREYRPDRSDYDIALVRLQGPEDCARFSSHVLPACLPMLRE 703
Db 401 LQLPWTECELSGYKHEALSPFYSERLKEAHVRLYPSSRCTSOHLNRTVTDNMLCAGD 460
704 RPKQTASNCYITIGWG--DT-GRAYSTRLOQAALPLPKRFC-ERYKGR-FTGRMLCAGN 758
Db 461 TRSGGQANLHADCGDSGGLVCLNDGAMTLV-GIISWGLCGGOKDVPGYTKVTNYLD 519

QY 759 LHE--HKKV-DSCGDSGGPLMCPGSGSVVYVTSWGTGCGVXDSPGVYTKVSAFVP 814
DB 520 WIR 522
11:
QY 815 WIR 817

DB 348 WIR 350
11:
QY 815 WIR 817

Search completed: Mon Mar 13 10:10:06 2000
Job time : 23 secs.

RESULT 7
ID US-08-427-640-6 STANDARD: PRT: 355 AA.
XX xxxxxx
XX
DT
XX

Sequence 6, Application US/08427640
Sequence 6, Application US/08427640
Patent No. 5658788

GENERAL INFORMATION:

CC APPLICANT: Berg et al.
CC TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disc
CC NUMBER OF SEQUENCES: 28
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Eli Lilly and Company
CC STREET: Lilly Corporate Center
CC CITY: Indianapolis
CC STATE: IN.
CC COUNTRY: U.S.A.
CC ZIP: 46285
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
CC COMPUTER: Macintosh
CC OPERATING SYSTEM: Macintosh
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/427,640
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/689,410
CC FILING DATE: 22 APRIL 1991
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 355 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 355 AA: 39563 MW: 638774 CN;

Seq Match 9.1%; Score 573; DB 1; Length 355;

Best Local Similarity 37.3%; Pred. No. 3,79e-41;
Matches 113; Conservative 66; Mismatches 102; Indels 22; Gaps 18;

DB 54 LGKHNCRNDGDAKPMCHVLRRLTWEXCDVPS-CSTGGLROYSQPOFRKGLPAD1 112
113 IGRHN-CRHSF-DAGVYCDVF-GKKAS-GNSNKSLSVYGLRLHRRQRIIGKNSLR 586
QY 531 ASHPWQAIRAKHRSGEPLGGLISSCWLTSAHCFQERF--PRHLLVTLGRYR 170
DB 113 ASHPWQAIRAKHRSGEPLGGLISSCWLTSAHCFQERF--PRHLLVTLGRYR 170
QY 587 GGPWQVSLKSSHDG-RLCGATLLSSCWLTSAHCFK-RYGNSTRYAVRG-DYH 643
DB 171 -VVGEEQKFEVEKYIVHKEFPDDTYNDIALQLKSDSRCAQESSVYRVCLPP-AD 228
QY 644 TLVPEFEFEELIGVQIVIHREYRPDSYDIALVRLQGPREGQCARSSHVLPACLPMLRE 703
DB 229 LQLPDTECLSGYKHEALSPYSERLKEAHVRLYPSSRCTSQHLNRTYDNLCAAD 288
QY 704 RPKTASNCITGWS-DI-GRAYSTRLOQAAPLPKRFK-DEYRKGR-FTGRMLCAGN 758
DB 289 TRSGGQANLHDCAGDSGGPLCNDGRNTLV-GIISWGLGCGKDPGVYTKVNYLD 347
QY 759 LHE--HKKV-DSCGDSGGPLMCPGSGSVVYVTSWGTGCGVXDSPGVYTKVSAFVP 814

THIS PAGE BLANK (USPTO)

 WISE (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 LPP protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:08:48 2000; MasPar time 32.43 seconds
 Tabular output not generated. 600,342 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (1-822) from US09147947A.pep (1 of 2)
 Perfect Score: 6328
 Sequence: 1 PTRRPPPLPRFPRLP.....PGVTKVSAPVFWIKSVTKL 822

Scoring table:
 PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 37.360; Variance 145.793; scale 0.256

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	6328	100.0	822	Human serine protease	0.00e+00
2	6328	100.0	875	Human neurotropsin.	0.00e+00
3	4199	66.4	761	Mouse neurotropsin.	0.00e+00
4	4199	66.4	761	Mouse serine protease	0.00e+00
5	781	12.3	1765	Human SRCR protein.	9.17e-68
6	729	11.5	347	Human scavenger recept	3.12e-62
7	729	11.5	347	Amino acid sequence of	3.12e-62
8	729	11.5	347	Human liver cell clone	3.12e-62
9	607	9.6	522	Delta (466-470) tPA va	4.02e-49
10	605	9.6	522	Delta (466-470) tPA va	4.02e-49
11	598	9.5	434	Delta (92-179) and delt	2.19e-48
12	599	9.5	522	Delta (466-470) tPA va	1.72e-48
13	598	9.5	522	Delta (466-470) tPA va	2.19e-48
14	600	9.5	559	JM1-229 cell line t-PA	1.35e-48
15	597	9.4	356	Amino acid sequence of	2.79e-48
16	597	9.4	522	Delta (466-470) tPA va	2.79e-48
17	590	9.3	390	Modified tissue plasmi	1.52e-47
18	585	9.3	395	Pro-urokinase derivati	2.46e-47
19	588	9.3	410	Pro-urokinase mutant F1	1.52e-47
20	590	9.3	522	Delta (466-470) tPA va	1.52e-47
21	584	9.2	378	Human prourokinase var	6.46e-47
22	584	9.2	389	Human prourokinase var	6.46e-47
23	584	9.2	401	Human prourokinase var	6.46e-47

ID	Query Match	Score	DB 1	Length	822	Best Local Similarity	100.0%	Pred. No.	0.00e+00	Matches	822	Conservative	0	Mismatches	0	Indels	0	Gaps	0
24	584	9.2	411	1	W13634	Human native prourokin	6.46e-47												
25	583	9.2	411	1	R47965	PUR N32P G38K G39K.	8.23e-47												
26	583	9.2	411	1	R47962	PUR G38A G39A.	8.23e-47												
27	583	9.2	411	1	R47966	PUR G16A.	8.23e-47												
28	583	9.2	411	1	R63003	Pro-urokinase mutant S	8.23e-47												
29	583	9.2	411	1	R47964	PUR N32P G38K.	8.23e-47												
30	583	9.2	411	1	R47959	PUR G32P.	8.23e-47												
31	583	9.2	411	1	R47961	PUR G38A.	8.23e-47												
32	583	9.2	411	1	R47957	PUR Y24A.	8.23e-47												
33	583	9.2	411	1	R47963	PUR G53A.	8.23e-47												
34	583	9.2	411	1	R40225	PUR.	8.23e-47												
35	583	9.2	411	1	R47960	PUR P34A.	8.23e-47												
36	583	9.2	411	1	R47958	PUR S26T.	8.23e-47												
37	583	9.2	411	1	R47956	PUR G16A G17A.	8.23e-47												
38	582	9.2	411	1	R63008	Pro-urokinase mutant S	1.05e-46												
39	582	9.2	411	1	R62999	Pro-urokinase mutant H	1.05e-46												
40	582	9.2	411	1	R63007	Pro-urokinase mutant S	1.05e-46												
41	582	9.2	411	1	P94410	Sequence of coding reg	8.23e-47												
42	582	9.2	525	1	R09230	t-PA deletion variant	1.05e-46												
43	584	9.2	526	1	R09227	t-PA deletion variant	6.46e-47												
44	583	9.2	527	1	R06236	Novel tissue plasminog	8.23e-47												
45	582	9.2	527	1	R09278	t-PA variant H432A, R4	1.05e-46												

ALIGNMENTS

RESULT 1
 ID W99087; standard; Protein: 822 AA.
 AC W99087;
 DT 13-MAY-1999 (first entry)
 DE Human serine protease BSSP-3.
 KW serine protease; BSSP-3; brain tissue.
 OS Homo sapiens.
 PN W09905280-A1.
 PD 04-FEB-1999.
 PF 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SUNTORY LTD.
 PI Tsuruoka N. Yamaguchi N. Yamashiro K;
 DR WPI: 99-142942/12.
 DR N-PSDB: X19024.
 PT New serine protease expressed in brain tissue - used in screening
 for potential serine protease inhibitors for drug use
 PS Claim 1; Page 61-65; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 is isolated from human brain tissue. Transformants may be used to
 produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 822 AA:

Query Match 100.0%; Score 6328; DB 1; Length 822;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	PTRRPPPLPRFPRLP	PAORPHALQGHTRPPHMC	CPAGEPVSTDFGAPLRMA	60
Qy	1	PTRRPPPLPRFPRLP	PAORPHALQGHTRPPHMC	CPAGEPVSTDFGAPLRMA	60
Db	61	EVPFPLERSPPASWAL	RGQRHNFCSR	PDGAGRPWCFTYDAGKGVWGYC	CDGRHGSVRLR
Qy	61	EVPFPLERSPPASWAL	RGQRHNFCSR	PDGAGRPWCFTYDAGKGVWGYC	CDGRHGSVRLR
Db	121	GKNEFEGETVEYASG	WCTVCS	SHWDS	DAVICHQLOL
Qy	121	GKNEFEGETVEYASG	WCTVCS	SHWDS	DAVICHQLOL
Db	181	WSNVRCRGDEENIL	LCERKIMOGGVC	POKMAAVTCSF	SHGTFPIRLAGSSVHEGRV
Qy	181	WSNVRCRGDEENIL	LCERKIMOGGVC	POKMAAVTCSF	SHGTFPIRLAGSSVHEGRV
Db	241	ELYHAGOMGTVCDD	WDADA	EVICROL	LSGIAKMAHVAE
Qy	241	ELYHAGOMGTVCDD	WDADA	EVICROL	LSGIAKMAHVAE

QY 241 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 300
 Db 301 LSTEOCKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVYRGOMGTVCD 360
 QY 301 LSTEOCKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVYRGOMGTVCD 360
 Db 361 DGTETLNTYYVVCQQLGKRYKQKASANHFESTGPIWLDVSCSEKRETRFLQCSRRQGRH 420
 QY 361 DGTETLNTYYVVCQQLGKRYKQKASANHFESTGPIWLDVSCSEKRETRFLQCSRRQGRH 420
 QY 361 DGTETLNTYYVVCQQLGKRYKQKASANHFESTGPIWLDVSCSEKRETRFLQCSRRQGRH 420
 Db 421 DCSHREDVSIACYPGEGGHLISLGFVPLMDGENKKEGRVVFINGOMGTICDDGWTDKD 480
 QY 421 DCSHREDVSIACYPGEGGHLISLGFVPLMDGENKKEGRVVFINGOMGTICDDGWTDKD 480
 Db 481 AAVICRQLGKGPAPARTMAVFEGBKGRPIHVDNVCYKCTGNERSLADCIKODIGRHNCHSE 540
 QY 481 AAVICRQLGKGPAPARTMAVFEGBKGRPIHVDNVCYKCTGNERSLADCIKODIGRHNCHSE 540
 Db 541 DAGVICTDYFEGKASGNSKESLSVCGRLRLHRQKRIIGKNSLRGWPQVSLRLKSS 600
 QY 541 DAGVICTDYFEGKASGNSKESLSVCGRLRLHRQKRIIGKNSLRGWPQVSLRLKSS 600
 Db 601 HGDGRLLCGATLLSSCWVLTAAHCFKRYGNSRSTSAVAVRGDYHTLVPEEPEEIEGVOQIV 660
 QY 601 HGDGRLLCGATLLSSCWVLTAAHCFKRYGNSRSTSAVAVRGDYHTLVPEEPEEIEGVOQIV 660
 Db 661 IHREYRPRSDYDIALVRLQGPEDOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 QY 661 IHREYRPRSDYDIALVRLQGPEDOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720
 Db 721 GRAYSTLQOAAIPLLPKRCCEERKGRFTGRMLCAGNLHEHKKRVDSOCGDSGGLMCR 780
 QY 721 GRAYSTLQOAAIPLLPKRCCEERKGRFTGRMLCAGNLHEHKKRVDSOCGDSGGLMCR 780
 Db 781 PGESWVYVGTSMGCGVKSPPGYTKVSAPWIKSVTKL 822
 QY 781 PGESWVYVGTSMGCGVKSPPGYTKVSAPWIKSVTKL 822

RESULT 2

ID W83361 standard; Protein: 875 AA.
 AC W83361:

DT 17-FEB-1999 (first entry)

DE Human neurotrophin; tumour inhibition; neurological disease;

KM lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KV epilepsy; hypoxia; ischaemia; nerve transection; neuroangiogenesis;

KX emptysema; bronchitis.

KY homo sapiens.

PS 09849322-A1.

24-APR-1998.

26-APR-1997: IB0625.

PA (SOUND) SONDREGER P.

PI Sondereger P.

DR WPI: 99-009438/01.

DR N-PSDB: V72589.

PT New human and murine neurotrophin - used, e.g. for inhibiting

PT tumours, treatment of neurological or lung disease, including by

PT gene therapy and in drug development

PS Claim 1: Page 20-24; 50pp: English.

CC The present sequence represents human neurotrophin. Neurotrophin proteins

CC and polynucleotides can be used: (i) to inhibit tumours, including

CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage

CC caused by stroke or brain injury (having a protective effect on the

CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

CC increase survival of damaged neurons (e.g. in cases of hypoxia,

CC ischaemia, nerve transection) and to stimulate regeneration and/or

CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

CC degeneration or neuroangiogenesis); (vi) to prevent apoptosis (or other

CC causes of cell death) in the nervous system; (vii) to regenerate brain

CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

CC performance, including learning and memory; (x) to treat or prevent a

CC wide range of psychiatric disorders; and (xi) to treat brain or lung

CC injury associated with protease expression (specifically emptysema or

CC bronchitis).

CC Sequence 875 AA:

QY 54 PTTREPPPLRFRPPRALPAORPHALQAGHTPRPPMCCPAGBEPVSVTDGACCLMA 113

QY 1 PTTREPPPLRFRPPRALPAORPHALQAGHTPRPPMCCPAGBEPVSVTDGACCLMA 60

Db 114 EVPPFLERSPASMAQLRGORHNFCSRPGAGPWCFCYDAGKQVMDGCDCRHGSVRLR 173

QY 61 EVPPFLERSPASMAQLRGORHNFCSRPGAGPWCFCYDAGKQVMDGCDCRHGSVRLR 120

Db 174 GGRNEEGIVEYASGVWGTSSHWDSDASYICHLQGLGKGLAKQTPFSGGLIPY 233

QY 121 GGRNEEGIVEYASGVWGTSSHWDSDASYICHLQGLGKGLAKQTPFSGGLIPY 180

Db 234 WSNVRCGDEENILCEKDIWOGVCPQKMAAVTCSFSGPTFPPIIRLAGSSVHEGRV 293

QY 181 WSNVRCGDEENILCEKDIWOGVCPQKMAAVTCSFSGPTFPPIIRLAGSSVHEGRV 240

Db 294 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 353

QY 241 ELYHAGOMGVCDQDDDAEVIICRQLGSLGAKAHQAVFEGSGSPVMDDEVCTGNE 300

Db 354 LSTEOCKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVYRGOMGTVCD 413

QY 301 LSTEOCKSSWGEHNCHEKEDAGVSCPTLDGVIIRLAGKSGHEGRLEVYRGOMGTVCD 360

Db 414 DGTETLNTYYVVCQQLGKRYKQKASANHFESTGPIWLDVSCSEKRETRFLQCSRRQGRH 473

QY 361 DGTETLNTYYVVCQQLGKRYKQKASANHFESTGPIWLDVSCSEKRETRFLQCSRRQGRH 420

Db 474 DCSHREDVSIACYPGEGGHLISLGFVPLMDGENKKEGRVVFINGOMGTICDDGWTDKD 533

QY 421 DCSHREDVSIACYPGEGGHLISLGFVPLMDGENKKEGRVVFINGOMGTICDDGWTDKD 480

Db 534 AAVICRQLGKGPAPARTMAVFEGBKGRPIHVDNVCYKCTGNERSLADCIKODIGRHNCHSE 593

QY 481 AAVICRQLGKGPAPARTMAVFEGBKGRPIHVDNVCYKCTGNERSLADCIKODIGRHNCHSE 540

Db 594 DAGVICTDYFEGKASGNSKESLSVCGRLRLHRQKRIIGKNSLRGWPQVSLRLKSS 653

QY 541 DAGVICTDYFEGKASGNSKESLSVCGRLRLHRQKRIIGKNSLRGWPQVSLRLKSS 600

Db 654 HGDGRLLCGATLLSSCWVLTAAHCFKRYGNSRSTSAVAVRGDYHTLVPEEPEEIEGVOQIV 713

QY 601 HGDGRLLCGATLLSSCWVLTAAHCFKRYGNSRSTSAVAVRGDYHTLVPEEPEEIEGVOQIV 660

Db 714 IHREYRPRSDYDIALVRLQGPEDOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 773

QY 661 IHREYRPRSDYDIALVRLQGPEDOCARFSSHVLPACPLMRERPOKTASNCYITGMDT 720

Db 774 GRAYSTLQOAAIPLLPKRCCEERKGRFTGRMLCAGNLHEHKKRVDSOCGDSGGLMCR 833

QY 721 GRAYSTLQOAAIPLLPKRCCEERKGRFTGRMLCAGNLHEHKKRVDSOCGDSGGLMCR 780

Db 834 PGESWVYVGTSMGCGVKSPPGYTKVSAPWIKSVTKL 875

QY 781 PGESWVYVGTSMGCGVKSPPGYTKVSAPWIKSVTKL 822

RESULT 3

ID W83362 standard; Protein: 761 AA.

AC W83362;

DT 17-FEB-1999 (first entry)

DE Mouse neurotrophin.

KM Mouse: neurotrophin; tumour inhibition; neurological disease;
 KM lung disease: gene therapy; drug development; stroke; brain injury;
 KM neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 KM epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;
 KM myopathy; bronchitis.
 OS Mus musculus.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998; I80625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR MPI; 99-009438/01.
 DR N-PSDB; V72390.
 PT New human and murine neurotrophin - used, e.g. for inhibiting
 tumours, treatment of neurological or lung disease, including by
 gene therapy and in drug development
 Claim 1: Page 29-32; 50pp; English.
 CC The present sequence represents mouse neurotrophin. Neurotrophin proteins
 and polynucleotides can be used: (i) to inhibit tumours, including
 metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 caused by stroke or brain injury (having a protective effect on the
 penumbra zone); (iii) to treat or prevent neurodegeneration,
 neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 increase survival of damaged neurons (e.g. in cases of hypoxia,
 ischaemia, nerve transection) and to stimulate regeneration and/or
 restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 degeneration or neurodegeneration); (vi) to prevent apoptosis (or other
 causes of cell death) in the nervous system; (vii) to regenerate brain
 and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 performance, including learning and memory; (x) to treat or prevent a
 wide range of psychiatric disorders; and (xi) to treat brain or lung
 injury associated with protease expression (specifically emphysema or
 bronchitis).
 CC Sequence 761 AA:

Query Match 66.4%; Score 4199; DB 1; Length 761;
 Best Local Similarity 87.0%; Pred. No. 0.00e+00;
 Matches 528; Conservative 51; Mismatches 28; Indels 0; Gaps 0;

DB 155 CDCGCGPALVITLVGNSGHEGVELLYAHAGWGTCDDQMDNADADVICRQLGSLGIK 214
 QY 216 CFSHSGPTPIRLAGSSVHEGVELLYAHAGWGTCDDQMDNADADVICRQLGSLGIK 275
 DB 215 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 274
 QY 276 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 335
 DB 275 LAGKSTHEGRLVLYKGGWGTVCDDGWTMTNTYVACRLGFFRYGKSSVNHFDGSRPI 334
 QY 336 LAGKSHGRLVLYKGGWGTVCDDGWTMTNTYVACRLGFFRYGKSSVNHFDGSRPI 395
 DB 335 WLDVSCSGKEVFFICSSRQWRGHDCHREDVGLTCYPSDSGHRSLSPGPPIRLVDGENK 394
 QY 396 WLDVSCSGKEVFFICSSRQWRGHDCHREDVGLTCYPSDSGHRSLSPGPPIRLVDGENK 455
 DB 395 KEGREVEFVNGWGTCDDGWTMDKHAAYICRQLGYGPARARTMAVFGGKGPITHDNYK 454
 QY 456 KEGREVEFVNGWGTCDDGWTMDKHAAYICRQLGYGPARARTMAVFGGKGPITHDNYK 515
 DB 455 CTGNEKALADCVKODIGRHNCRHSEDAGVICYLEKKAASSGKEMLSGCGRLHLHRQ 514
 QY 516 CTGNEKALADCVKODIGRHNCRHSEDAGVICYLEKKAASSGKEMLSGCGRLHLHRQ 575
 DB 515 KRITGNNLSRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRGKNSRST 574
 QY 576 KRITGNNLSRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRGKNSRST 635
 DB 575 AVAVGVDHTLVPEEPEEIGVOOIVIHRYNRPDRSDIDIALVRLQGPGECAFFSSHVLP 634
 QY 636 AVAVGVDHTLVPEEPEEIGVOOIVIHRYNRPDRSDIDIALVRLQGPGECAFFSSHVLP 695
 DB 635 ACPLWREPRQKATSNCHITGWGDTGRAVSRTLOQAAPLLPRFCRKERYKGLFTGMLC 694

QY 696 ACPLWREPRQKATSNCHITGWGDTGRAVSRTLOQAAPLLPRFCRKERYKGLFTGMLC 755
 DB 695 AGNLQEDNRVDSGCGSGGLMCKEPPDESVMVYGVTSWGYCGCVKDPGYTVFVAPV 754
 QY 756 AGNLQEDNRVDSGCGSGGLMCKEPPDESVMVYGVTSWGYCGCVKDPGYTVFVAPV 815
 DB 755 IKSVTSL 761
 QY 816 IKSVTSL 822

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088;
 DE 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 KM Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN W09905290-A1.
 PF 04-FEB-1999.
 PF 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SUNTORX LTD.
 PI Tsutoka N, Yamaguchi N, Yamashiro K;
 DR N-PSDB; X18027.
 PT New serine protease expressed in brain tissue - used in screening
 for potential serine protease inhibitors for drug use
 Example 1: Page 51-54; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 is isolated from mouse brain tissue. Transformants may be used to
 produce the enzyme or its partial sequences. Products from the present
 invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 CC Sequence 761 AA:

Query Match 66.4%; Score 4199; DB 1; Length 761;
 Best Local Similarity 87.0%; Pred. No. 0.00e+00;
 Matches 528; Conservative 51; Mismatches 28; Indels 0; Gaps 0;

DB 155 CDCGCGPALVITLVGNSGHEGVELLYAHAGWGTCDDQMDNADADVICRQLGSLGIK 214
 QY 216 CFSHSGPTPIRLAGSSVHEGVELLYAHAGWGTCDDQMDNADADVICRQLGSLGIK 275
 DB 215 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 274
 QY 276 AMHQAHEGSGPILDEVRCTGNELSIEQCPSSWGEHNCCKEDAGVSCVPLTDGVR 335
 DB 275 LAGKSTHEGRLVLYKGGWGTVCDDGWTMTNTYVACRLGFFRYGKSSVNHFDGSRPI 334
 QY 336 LAGKSHGRLVLYKGGWGTVCDDGWTMTNTYVACRLGFFRYGKSSVNHFDGSRPI 395
 DB 335 WLDVSCSGKEVFFICSSRQWRGHDCHREDVGLTCYPSDSGHRSLSPGPPIRLVDGENK 394
 QY 396 WLDVSCSGKEVFFICSSRQWRGHDCHREDVGLTCYPSDSGHRSLSPGPPIRLVDGENK 455
 DB 395 KEGREVEFVNGWGTCDDGWTMDKHAAYICRQLGYGPARARTMAVFGGKGPITHDNYK 454
 QY 456 KEGREVEFVNGWGTCDDGWTMDKHAAYICRQLGYGPARARTMAVFGGKGPITHDNYK 515
 DB 455 CTGNEKALADCVKODIGRHNCRHSEDAGVICYLEKKAASSGKEMLSGCGRLHLHRQ 514
 QY 516 CTGNEKALADCVKODIGRHNCRHSEDAGVICYLEKKAASSGKEMLSGCGRLHLHRQ 575
 DB 515 KRITGNNLSRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRGKNSRST 574
 QY 576 KRITGNNLSRGAMPQASRLRSAGDGRLLCGATLLSSCWLTAAHCFKRGKNSRST 635
 DB 575 AVAVGVDHTLVPEEPEEIGVOOIVIHRYNRPDRSDIDIALVRLQGPGECAFFSSHVLP 634
 QY 636 AVAVGVDHTLVPEEPEEIGVOOIVIHRYNRPDRSDIDIALVRLQGPGECAFFSSHVLP 695

Db 84 EKEQKVLIOVSCTGTEPTDLAOCDEE--VYDCSHDEPDAGASCENPESSEFSYPGVR1A 141
 QY 284 EGGPVMLEDEVRCTGNETLIEQCPKSSWGEHNGHEDAGVSC-TPLETD-G-V--IRLA 337
 Db 142 DGGHCKGRVEYKHONQWYVTCOTGMSLRAKVYVROLCGSAVLTQKCNKHAYGRKPI 201
 QY 338 GGGSHGRLLEYVYRGQKGTVCDDGTELTNTYVVCRLG-FKYG-KQASANFEESTGPI 395
 Db 202 WLSQMSCSGREATLDDCPSPGPKNTCNHDEDTWEC---ED---P--FDLRLVGGDNL 252
 QY 396 WLDVSCSGKRETRFLQCSRWGRHDCSHREDVSIACYPGGEGHRLSLGFPVRLMDGENK 455
 Db 253 CGSRLEVLHKGWGSVCDNMGKEQDYVCKQJGCKSLSPSPFRDKCYGPGVGRITWLN 312
 QY 456 KEGRVEVFPNGQGTICDDGWTDKDAVAICRLQGY-KGPARA-RIMAYFGEGKGP1HYVN 513
 313 VRCSGEOSLQCOHRFMGFHCTHOEDVAVIC 345
 514 VKCTGNERSLADCIKODIGRHNCRHSEDAVAVIC 546

RESULT 7
 ID Y13369 standard; Protein: 347 AA.
 AC Y13369:
 DI 25-JUN-1999 (first entry)
 DE Amino acid sequence of protein PRO229.
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 PN W09914328-A2.
 PD 25-MAR-1999.
 PF 16-SEP-1996; U19330.
 PR 25-NOV-1997; US-066840.
 PR 17-SEP-1997; US-059113.
 PR 17-SEP-1997; US-059115.
 PR 17-SEP-1997; US-059117.
 PR 17-SEP-1997; US-059119.
 PR 17-SEP-1997; US-059121.
 PR 17-SEP-1997; US-059122.
 PR 17-SEP-1997; US-059184.
 PR 18-SEP-1997; US-059263.
 PR 18-SEP-1997; US-059266.
 PR 15-OCT-1997; US-062125.
 PR 17-OCT-1997; US-062285.
 PR 17-OCT-1997; US-062287.
 PR 21-OCT-1997; US-063486.
 PR 24-OCT-1997; US-062814.
 PR 24-OCT-1997; US-062816.
 PR 24-OCT-1997; US-063045.
 PR 24-OCT-1997; US-063120.
 PR 24-OCT-1997; US-063121.
 PR 24-OCT-1997; US-063127.
 PR 24-OCT-1997; US-063128.
 PR 27-OCT-1997; US-063329.
 PR 27-OCT-1997; US-063327.
 PR 28-OCT-1997; US-063541.
 PR 28-OCT-1997; US-063542.
 PR 28-OCT-1997; US-063544.
 PR 28-OCT-1997; US-063549.
 PR 28-OCT-1997; US-063550.
 PR 28-OCT-1997; US-063564.
 PR 29-OCT-1997; US-063435.
 PR 29-OCT-1997; US-063704.
 PR 29-OCT-1997; US-063732.
 PR 29-OCT-1997; US-063738.
 PR 29-OCT-1997; US-063734.
 PR 29-OCT-1997; US-064215.
 PR 29-OCT-1997; US-063735.

PR 31-OCT-1997; US-063870.
 PR 31-OCT-1997; US-064103.
 PR 03-NOV-1997; US-064248.
 PR 07-NOV-1997; US-064809.
 PR 12-NOV-1997; US-065186.
 PR 17-NOV-1997; US-065846.
 PR 18-NOV-1997; US-065893.
 PR 21-NOV-1997; US-066120.
 PR 21-NOV-1997; US-066364.
 PR 24-NOV-1997; US-066772.
 PR 24-NOV-1997; US-066466.
 PR 24-NOV-1997; US-066770.
 PR 24-NOV-1997; US-066511.
 PR 24-NOV-1997; US-066453.
 PA (Genentech) GENENTECH INC.
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 DR WPI: 98-229533/19.
 DR N-PSDB: X52240.
 PT New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PS Claim 12; Fig 54, 320pp; English.
 CC Y13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to growth
 CC or survival of nerve cells including Parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata. PRO266 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair. PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 SQ Sequence 347 AA;

Query Match 11.5%; Score 729; DB 1; Length 347;
 Best Local Similarity 40.5%; Pred No. 3 12e-62;
 Matches 135; Conservative 64; Mismatches 110; Indels 24; Gaps 14;

Db 24 VRLVGLHRCRGEVNEVEQKQWGTVCDDGWDIKDAVAICRELDCGASGTPGILYEPRA 83
 QY 227 IRLAGSSVHEGRVELYHAGWGTVCDDGWDADAEVLCRQYG-LSGIKAMHQA-Y-FG 283
 Db 84 EKEQKVLIOVSCTGTEPTDLAOCDEE--VYDCSHDEPDAGASCENPESSEFSYPGVR1A 141
 QY 284 EGGPVMLEDEVRCTGNETLIEQCPKSSWGEHNGHEDAGVSC-TPLETD-G-V--IRLA 337
 Db 142 DGGHCKGRVEYKHONQWYVTCOTGMSLRAKVYVROLCGSAVLTQKCNKHAYGRKPI 201
 QY 338 GGGSHGRLLEYVYRGQKGTVCDDGTELTNTYVVCRLG-FKYG-KQASANFEESTGPI 395
 Db 202 WLSQMSCSGREATLDDCPSPGPKNTCNHDEDTWEC---ED---P--FDLRLVGGDNL 252
 QY 396 WLDVSCSGKRETRFLQCSRWGRHDCSHREDVSIACYPGGEGHRLSLGFPVRLMDGENK 455
 Db 253 CGSRLEVLHKGWGSVCDNMGKEQDYVCKQJGCKSLSPSPFRDKCYGPGVGRITWLN 312
 QY 456 KEGRVEVFPNGQGTICDDGWTDKDAVAICRLQGY-KGPARA-RIMAYFGEGKGP1HYVN 513
 Db 313 VRCSGEOSLQCOHRFMGFHCTHOEDVAVIC 345
 QY 514 VKCTGNERSLADCIKODIGRHNCRHSEDAVAVIC 546

Thu Mar 16 07:57:15 2000

US-09-147-947-6-01.rag

Page 6

Search completed: Mon Mar 13 10:09:25 2000
Job time : 37 secs.

QY 276 AMHQAAYEGGSGPVMDEVRCTGNETLSEOCPSKSNMGEHNGKEDAGVSTPLTDEVIR 335
Db 275 LAGKASTHESRLVYKGGMGTYCDDGTEMTYVACRLLEGFKYKQSSVNHDEGSNRP 334
QY 336 LAGKASHESRLVYKGGMGTYCDDGTEMTYVACRLLEGFKYKQSSVNHDEGSNRP 335
Db 335 WLDVSCSGKEVEFICOSRQWGRHDCSHREDVLTCTPPSDGRLSPGPRIYVDENK 394
QY 396 WLDVSCSGKEVEFICOSRQWGRHDCSHREDVLTCTPPSDGRLSPGPRIYVDENK 455
Db 395 KEGREVEFVNGWGCTICDDGTEMTKHAAYICROLGYKGPAPARTMAYEGEGKPIHMDNVK 454
QY 456 KEGREVEFVNGWGCTICDDGTEMTKHAAYICROLGYKGPAPARTMAYEGEGKPIHMDNVK 515
Db 455 CTGNEKALADCVKODIGRHNCRHSEDAVICTYLEKRASSSGNEMLSGGCGLLHRRQ 514
QY 516 CTGNEKSLADCIKODIGRHNCRHSEDAVICTYLEKRASSSGNEMLSGGCGLLHRRQ 575
Db 515 KRIGGNNSLRGAMPQASRLRSAGHGRLLCGATILLSGCWLTAAHCKRKYNNRSY 574
QY 576 KRIGGNNSLRGAMPQASRLRSAGHGRLLCGATILLSGCWLTAAHCKRKYNNRSY 635
Db 575 AAVRGYHTLYPEEFOEIGVQIVIHRYNRPDRSDIALVRLQEGEOCARLSTHVP 634
QY 636 AAVRGYHTLYPEEFOEIGVQIVIHRYNRPDRSDIALVRLQEGEOCARLSTHVP 695
Db 635 ACPLMRERPOKTASNCITGWGDTGRAYSRTLOQAAVPLLPKRECKERYKGLFTGMCL 694
QY 696 ACPLMRERPOKTASNCITGWGDTGRAYSRTLOQAAVPLLPKRECKERYKGLFTGMCL 755
Db 695 AGLOEDNVDSOGSGGGLMCEKPEDESVMVYGVTSWGGCGKDPGYTRAPAPVP 754
QY 756 AGLOEDNVDSOGSGGGLMCEKPEDESVMVYGVTSWGGCGKDPGYTRAPAPVP 815
Db 755 IKSVTSU 761
QY 816 IKSVTKL 822

RESULT 2
ENTRY 114893 #type complete
TITLE scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name purple urchin
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSION 114893
REFERENCE 218253
AUTHORS Pancer, Z.; Rast, J.P.; Davidson, E.H.
JOURNAL Immunogenetics (1999) 49:773-786
TITLE Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.
#accession 114893
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2153 #label PAN
#cross-references EMBL:AF064259; NID:g4165052; PID:g4165053; PID:AD08654.1

GENETICS
#gene SRCR12
SUMMARY #length 2153 #molecular-weight 226556 #checksum 3510

Query Match 21.6%; Score 1370; DB 2; Length 2153;
Best Local Similarity 44.5%; Pred. No. 0.00e+00;
Matches 209; Conservative 99; Mismatches 143; Indels 19; Gaps 13;

Db 433 C-PHNGVSHNCAHGEDAG-VSCA-PSSQESRYRLVGLNNREGRAVEIFLNNMGTYCDD 489
QY 85 CRSPDAGRPWCFYGDARGVDMWGYCDCHRGVRLRGKNEFEGYVEYASGVWGYCSS 144
Db 490 DMGTDPANVVCOLGIFSGGSARSSAYFGKSGPILLDNVCGSGNENSLCLCSNNGICVH 549

QY 145 HMDSDBASVICHQDLOGGKIAKOTFPFSLGLPIYMSVNRCKDEENLLEKIMWGG 204
Db 550 NCGHOEDASVYCTGYTTPMKNL-RIVGGSPSEGEVYVGHNGTYCDDMLINDANY 608
QY 205 VCPQKMAAVTCS-FSHGPTFPIIRLAGGSSVHEGVELYHAGMGTYCDDQMDADAEV 263
Db 609 VCRELGSAAITSSASGQSGDILLDDLRCSGTESLLTCPRHGVNHCANSEADAG 668
QY 264 ICROGLISGLAKMHOAYTGEESGPMVLDEVRCTGNETLSEOCPSKSNMGEHNGKEDAG 323
Db 669 VCASTSGPVGTMYRLVGGQNSRGRLEISINMGVTCDDMDINATVACQOLGS 728
QY 324 VSCPTPLDGV-RLAGKASHESRLVYKGGMGTYCDDGTEMTYVACRLLEGFKY 378
Db 729 SAVAPTSAHFGGSGTLMWLDVSCAGNENSLMDCGHRGLGVNCAHAEADAGVCIAS-D 787
QY 379 YG-KQASANHFEESTGPIWLDVSCGKETRFLQSRQWGRHDCSHREDVSIACYPGE 437
Db 788 G---PLN--IRLAGRSGHEGRVEISLGGDWGTCDDSGIEDAHVYCRQLGF-GPALSA 841
QY 438 GHRSLGFPVRLMDGKNEKREGEVEFVINGMGTYCDDGTEMTKHAAYICROLGYKGP 496
Db 842 VTASFGGSGSLMNVQSGDEATIAECSHNGICSHNGCHQEDAGVYC 891
QY 497 RTMAYFEGEGKPIHVDNKTGNERSLADCIKODIGRHNCRHSEDAVICT 546

RESULT 3
ENTRY 536077 #type complete
TITLE M130 antigen human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Aug-1996
ACCESSION 138003; 536077
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
TITLE A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#accession 138003
#status preliminary: translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1116 #label RES
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147
151-254 #domain scavenger receptor cysteine-rich domain homology
258-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology
470-573 #domain scavenger receptor cysteine-rich domain homology
575-678 #domain scavenger receptor cysteine-rich domain homology
711-814 #domain scavenger receptor cysteine-rich domain homology
816-920 #domain scavenger receptor cysteine-rich domain homology
921-1024 #domain scavenger receptor cysteine-rich domain homology
SUMMARY #length 1116 #molecular-weight 120979 #checksum 5687

Query Match 14.9%; Score 946; DB 2; Length 1116;
Best Local Similarity 39.2%; Pred. No. 6.78e-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSCGRVYKVOEEMGTVCNNGSMENVSYICNOLG-CPTAIPAKPGMANSAG 104
QY 117 VRLGKGNKEFEGETVEYVASGVWGTVCSHWDSDASVICHOLGKGKIAKOT-PFSGIG 175
Db 105 SGRI-WMDHVSRCNGESALMDCKHDGKHSNCTHOODAGVTC--DGSNLEM-RLRGG 160
QY 176 LIPIY-SNVRCRDEENILLCEKDIW-QGVCOPQKMAAAVTCFSHGPTPIRLAGGS 233
Db 161 NMSGRIEIKFO-GRMGTVCDNDFNIDHASYICROLE-CGSVAFSSSSNFGESGPIWF 218
QY 234 SYHGRVEL-YHAGOMGTVCDDQWDADAEVICROLGSLGIAKAMH-QATFGESSGPVML 291
Db 219 DDLICNGESALMCKHQGKHNCHDAEDAGVYCSKADLSRLVDGVECSGRLEVR 278
QY 292 DEVACTGNELIEQCPKSSWGEHNCHEKEDAGVCTPLTDGIVIRLAGKSGSHEGRLEVVY 351
Db 279 OGEWGTICDDGMDSYDAVAACOLGCPATAIAGRVAASKGFHIMLDSVSCGHEPAVW 338
QY 352 RGOMGTVCDDGMDLNTYVVCROLGFRYKQASAN-HFEESTGPIWDDVSCSGKETRFL 410
Db 339 QCKHEWGHKHYCNHNEEDAGVTC-SDG-SD-DEL---RLRGGSRGAGTVEVEIORLLGK 391
QY 411 QCSRQMGHRHDCSHREDVSIACYPGEGHRLSLGFPVRLMDGKKGKRGVETFINQOMGT 470
Db 392 VCDRGKGLKADVYVCROLGCGSALKTSYQVYSKIQATNTWLFSSCNGNETSLMDCKNMQ 451
QY 471 ICDDGMDTKDAVAICROLGYKGPARTATMAYFEGKGPFIHVNDVKTGNERSLADCIKOD 530
Db 452 WGGTICDHYEAKITC 467
QY 531 IGRHNCRHSDEAGVTC 546

RESULT 4
ENTRY 138006 #type complete
TITLE M130 antigen (extracellular variant) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 02-Aug-1996
ACCESSIONS I38006; S35768
REFERENCE I38003
#authors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of
#the scavenger receptor superfamily.
#cross-references MIMD:93380506
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1149 #label RES
#cross-references EMBL:222971; NID:9312148
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147
151-254 #domain scavenger receptor cysteine-rich domain homology
258-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology
470-573 #domain scavenger receptor cysteine-rich domain homology
608-711 #domain scavenger receptor cysteine-rich domain homology
744-847 #domain scavenger receptor cysteine-rich domain homology
849-953 #domain scavenger receptor cysteine-rich domain homology
954-1057 #domain scavenger receptor cysteine-rich domain homology
SUMMARY #length 1149 #molecular_weight 124328 #checksum 487

Query Match 14.9%; Score 946; DB 2; Length 1149;
Best Local Similarity 39.2%; Pred. No. 6,78e-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSCGRVYKVOEEMGTVCNNGSMENVSYICNOLG-CPTAIPAKPGMANSAG 104
QY 117 VRLGKGNKEFEGETVEYVASGVWGTVCSHWDSDASVICHOLGKGKIAKOT-PFSGIG 175
Db 105 SGRI-WMDHVSRCNGESALMDCKHDGKHSNCTHOODAGVTC--DGSNLEM-RLRGG 160
QY 176 LIPIY-SNVRCRDEENILLCEKDIW-QGVCOPQKMAAAVTCFSHGPTPIRLAGGS 233
Db 161 NMSGRIEIKFO-GRMGTVCDNDFNIDHASYICROLE-CGSVAFSSSSNFGESGPIWF 218
QY 234 SYHGRVEL-YHAGOMGTVCDDQWDADAEVICROLGSLGIAKAMH-QATFGESSGPVML 291
Db 219 DDLICNGESALMCKHQGKHNCHDAEDAGVYCSKADLSRLVDGVECSGRLEVR 278
QY 292 DEVACTGNELIEQCPKSSWGEHNCHEKEDAGVCTPLTDGIVIRLAGKSGSHEGRLEVVY 351
Db 279 OGEWGTICDDGMDSYDAVAACOLGCPATAIAGRVAASKGFHIMLDSVSCGHEPAVW 338
QY 352 RGOMGTVCDDGMDLNTYVVCROLGFRYKQASAN-HFEESTGPIWDDVSCSGKETRFL 410
Db 339 QCKHEWGHKHYCNHNEEDAGVTC-SDG-SD-DEL---RLRGGSRGAGTVEVEIORLLGK 391
QY 411 QCSRQMGHRHDCSHREDVSIACYPGEGHRLSLGFPVRLMDGKKGKRGVETFINQOMGT 470
Db 392 VCDRGKGLKADVYVCROLGCGSALKTSYQVYSKIQATNTWLFSSCNGNETSLMDCKNMQ 451
QY 471 ICDDGMDTKDAVAICROLGYKGPARTATMAYFEGKGPFIHVNDVKTGNERSLADCIKOD 530
Db 452 WGGTICDHYEAKITC 467
QY 531 IGRHNCRHSDEAGVTC 546

RESULT 5
ENTRY 138004 #type complete
TITLE M130 antigen (cytosolic variant 1) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997
ACCESSIONS I38004; S36078
REFERENCE I38003
#authors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of
#the scavenger receptor superfamily.
#cross-references MIMD:93380506
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1151 #label RES
#cross-references EMBL:222969; NID:9312143; PID:9312144
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147
151-254 #domain scavenger receptor cysteine-rich domain homology
258-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology
470-573 #domain scavenger receptor cysteine-rich domain homology
575-678 #domain scavenger receptor cysteine-rich domain homology
711-814 #domain scavenger receptor cysteine-rich domain homology

816-920 #label SRC7\
#domain scavenger receptor cysteine-rich domain homology
#label SRC8\
921-1024 #domain scavenger receptor cysteine-rich domain homology
#label SRC9
#length 1151 #molecular-weight 124820 #checksum 1481

Query Match 14.9%; Score 946; DB 2; Length 1151;
Best Local Similarity 39.2%; Pred. No. 6,786-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSGRYEVQVEENGTVCNNGSMENAVSYICNOLG-CPTAIKAPGMANSSAG 104
QY 117 VRLRGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHQLQGGKIAKOT-PFSGLG 175
Db 105 SGRI-WMDHYSCRGNEBALMDCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
QY 176 LPIIYW-SNVRKRDENILCEKDIW-QGVCVPQKMAAAVTCFSHGPTPIIRLAGGS 233
QY 161 NMCSGRLEIKFO-GRWGTVCDDNFNIDHASYICROLE-CGSAYFSFGSSNFGSGSPIWF 218
234 SYHEGRVEL-YHAGQWGTVCDDQWDDADAEVICRQLGSLGIKAMW-QAYFGSGSPVYL 291
Db 219 DDILCNGNEBALNMCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 278
QY 292 DEVRCTGNEELISDCPRSSWGEHNGKHEKEDAGVSCPTLDVIRLAGGSGHEGRLEYY 351
Db 279 QGEMGTICDDGMPDSYDAVAVACOLGCPATAVTAIGRVNASKGFHIMLDSVSCOGHEPAW 338
QY 352 RGMWGTVCDDGWTLENTIYVVCROLGFKYKQASAN-HFEESTGPIWDDVSCSGKETREL 410
Db 339 QCKHHEWGHKHYCNHEDAGVTC-SDG-SD-LEL---RLRGGSGRCAGTVEVEIQRLLGK 391
QY 411 QCSRQWGRHDCSHREDVSIACYPGEGHRLSLGFVRLMDGENKKEGVEVFIINGQWGT 470
Db 392 VCDRGWGLKADVYVCRQLGCGSALKTSYQYYSKIQTATNWLFLSSCNGENELMDCKRMQ 451
QY 471 ICDDGWTCKDAAYICROLGKGPARTMAVFGEGKPIHVDNVKCTGNEBSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRNCRHSEDAGVTC 546

RESULT 6
ENTRY I38005 #type complete
TITLE M130 antigen (cytosolic variant 2) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
07-Feb-1997

ABSTRACTS
AUTHORS I38005; S36079
I38003
Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
Willis, A.C.; Mason, D.Y.
Eur. J. Immunol. (1993) 23:2320-2325
A new macrophage differentiation antigen which is a member of
the scavenger receptor superfamily.
#cross-references NUID:93380506

#accession I38005
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1156 ##label RES
#cross-references EMBL:222970; NID:9312145; PID:9312146
CLASSIFICATION
KEYWORDS
cytosol

FEATURE
43-147 #domain scavenger receptor cysteine-rich domain homology
151-254 #domain scavenger receptor cysteine-rich domain homology
256-361 #domain scavenger receptor cysteine-rich domain homology
365-468 #domain scavenger receptor cysteine-rich domain homology

470-573 #label SRC4\
#domain scavenger receptor cysteine-rich domain homology
#label SRC5\
575-678 #domain scavenger receptor cysteine-rich domain homology
#label SRC6\
711-814 #domain scavenger receptor cysteine-rich domain homology
#label SRC7\
816-920 #domain scavenger receptor cysteine-rich domain homology
#label SRC8\
921-1024 #domain scavenger receptor cysteine-rich domain homology
#label SRC9
#length 1156 #molecular-weight 125352 #checksum 9043

Query Match 14.9%; Score 946; DB 2; Length 1156;
Best Local Similarity 39.2%; Pred. No. 6,786-199;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVDGENKSGRYEVQVEENGTVCNNGSMENAVSYICNOLG-CPTAIKAPGMANSSAG 104
QY 117 VRLRGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHQLQGGKIAKOT-PFSGLG 175
Db 105 SGRI-WMDHYSCRGNEBALMDCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
QY 176 LPIIYW-SNVRKRDENILCEKDIW-QGVCVPQKMAAAVTCFSHGPTPIIRLAGGS 233
QY 161 NMCSGRLEIKFO-GRWGTVCDDNFNIDHASYICROLE-CGSAYFSFGSSNFGSGSPIWF 218
234 SYHEGRVEL-YHAGQWGTVCDDQWDDADAEVICRQLGSLGIKAMW-QAYFGSGSPVYL 291
Db 219 DDILCNGNEBALNMCKHDKGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 278
QY 292 DEVRCTGNEELISDCPRSSWGEHNGKHEKEDAGVSCPTLDVIRLAGGSGHEGRLEYY 351
Db 279 QGEMGTICDDGMPDSYDAVAVACOLGCPATAVTAIGRVNASKGFHIMLDSVSCOGHEPAW 338
QY 352 RGMWGTVCDDGWTLENTIYVVCROLGFKYKQASAN-HFEESTGPIWDDVSCSGKETREL 410
Db 339 QCKHHEWGHKHYCNHEDAGVTC-SDG-SD-LEL---RLRGGSGRCAGTVEVEIQRLLGK 391
QY 411 QCSRQWGRHDCSHREDVSIACYPGEGHRLSLGFVRLMDGENKKEGVEVFIINGQWGT 470
Db 392 VCDRGWGLKADVYVCRQLGCGSALKTSYQYYSKIQTATNWLFLSSCNGENELMDCKRMQ 451
QY 471 ICDDGWTCKDAAYICROLGKGPARTMAVFGEGKPIHVDNVKCTGNEBSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRNCRHSEDAGVTC 546

RESULT 7
ENTRY A32751 #type complete
TITLE spectrat receptor precursor - sea urchin (Strongylocentrotus
purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name
purple urchin
DATE 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change
02-Aug-1996
ACCESSIONS A32751; A31267
REFERENCE A32751
#authors Dangott, L.J.; Jordan, J.E.; Bellet, R.A.; Garbers, D.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:2128-2132
#title Cloning of the mRNA for the protein that crosslinks to the
egg peptide spectrat.
#cross-references NUID:89184581

FEATURE
#accession A32751
#molecule_type mRNA
#residues 1-352,'G',354-532 ##label DA2
#cross-references GB:J04518
##note the authors translated the codon CAC for residue 353 as
Gly
##note part of this sequence was confirmed by protein
sequencing

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS membrane protein

FEATURE

1-30 #domain signal sequence #status predicted #label SIG\
31-532 #product sceract receptor #status predicted #label MAT\
40-144 #domain scavenger receptor cysteine-rich domain homology
150-257 #label SRC1\
#domain scavenger receptor cysteine-rich domain homology
261-366 #label SRC2\
#domain scavenger receptor cysteine-rich domain homology
379-485 #label SRC3\
#domain scavenger receptor cysteine-rich domain homology

SUMMARY #length 532 #molecular-weight 57820 #checksum 5701

Very Match 13.5%; Score 852; DB 2; Length 532;

Local Similarity 33.5%; Pred. No. 4,36e-175; Mismatches 177; Indels 24; Gaps 23;

Conservative 99; Mismatches 177; Indels 24; Gaps 23;

DB 41 GNIRLIHGRTENEGSVEIYHATRWGVCWMMHNNANTCKQLGPPGARQFYRAYF-G 99
115 GSVRLRGKNEFEGTVEYASVGWGTCSHMDSDASVICHQLQL-GKGIKQTPFSG 173
DB 100 -AHVTFVYKMKNCIGNETLEDCYHRRPYRPMICNQAQAGVCELPKDEPGSL-RML 157
174 LGIPIYWS-NVRCRGDEENILLC-EKDIWGGVCPOKMAAAVTCFSHGPTEPIIRLAG 231
DB 158 GDVPMNEGTLTFWDGAMGSVCHTDFTPDGNVACROWGYSRGVKSITKDGHPGSTGPII 217
232 GSSVHEGRVELYHAGOWGTVCDDQMDDAEYICROLGLS-GIAKAMHQAYPEGSGPVM 290
DB 218 LDAVDEGTEAHITTECNMPTPYOACPYTHNMDVGVCKPNEGDIRLMDGSGPHGRY 277
291 LDEVRCGNELSTEQCKSSWG-EHNGHK-E-DAGVCTPLTDGYIRLAGKSHGRLL 347
DB 278 EIVHDDAMGTICDDGMDADANVYCRQAGYRGAVKASGEKEDFGFTWAPIHTSFVWCTG 337
348 EYVYRGOMGTVCDDGTETLNTYVVCROLGFKYGRKQASNHFE-S-T-GPIWLDVSCSG 404
DB 338 VEDRLIDCILRDGMT-HSCYHVEDASVVC-ATDDDTIELPKHTIRYRIVMGOGGRVE 395
405 KETRFLOCSRRQ-WGRHDCSHREDVSTACYPGEGHRLSLGF-PVRL-MDGENKKKEGRVE 461
DB 396 VSLGNGWRVCDPDSDEAKTYCYHAGYKMGASRAAGSAEVSAPFDLEAPFIIDGITCS 455
462 VFINGQWGTICDDGMDKDAVAVICROLGYK-GPARATMAYFGE--G-KGPIHVDNVCYK 517
DB 456 GVENETISOCOMKVSADMTCA-TGDVGVCCE 485
518 GNER-SIADCIKODIGHNCRHSESDAGVID 547

Search completed: Mon Mar 13 10:08:30 2000
Job time : 55 secs.

THIS PAGE BLANK (USPTO)

M O S E R
(TM)

Release 3.1a John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:04:57 2000; MasPar time 33.12 Seconds
Tabular output not generated. 741.229 Million cell updates/sec

Title: >US-09-147-947-6
Description: (1-822) from US09147947A.pep (1 of 2)
Perfect Score: 6328
Sequence: 1 PTRPPPLPRFPFRPPRALP.....PGVYTKVSAFVPIKSVTKL 822

Scoring table: PAM 150
Gap 11

Searched: 82229 segs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 50.140; Variance 71.058; scale 0.706

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length DB	ID	Description	Pred. No.
1	6328	100.0	875	1 NETR_HUMAN NEUTROTYPIN PRECURSOR	0.00e+00
2	4199	66.4	761	1 NETR_MOUSE NEUTROTYPIN PRECURSOR	0.00e+00
3	852	13.5	532	1 SPER_STRPU EGG PEPTIDE SPEPACT RE	2.16e-200
4	838	13.2	1436	1 WC11_BOVIN ANTIGEN WC1.1	2.38e-196
5	634	10.0	625	1 FA11_HUMAN COAGULATION FACTOR XI	5.45e-158
6	607	9.6	625	1 UROT_HUMAN TISSUE PLASMINOGEN ACT	2.30e-130
7	609	9.6	638	1 KAL_MOUSE PLASMA KALLIKREIN PREC	6.28e-131
8	601	9.5	415	1 ACRO_PIG ACROSIN PRECURSOR (EC	1.13e-118
9	600	9.5	559	1 UROT_RAT TISSUE PLASMINOGEN ACT	2.16e-128
10	603	9.5	638	1 KAL_RAT PLASMA KALLIKREIN PREC	3.09e-129
11	603	9.5	638	1 KAL_MOUSE PLASMA KALLIKREIN PREC	3.09e-129
12	580	9.2	431	1 ACRO_RABIT ACROSIN PRECURSOR (EC	9.09e-123
13	580	9.2	431	1 UROK_HUMAN UROKINASE-TYPE PLASMIN	9.09e-123
14	583	9.2	431	1 ACRO_RAT ACROSIN PRECURSOR (EC	1.31e-113
15	576	9.2	559	1 UROT_MOUSE TISSUE PLASMINOGEN ACT	3.58e-124
16	576	9.1	436	1 ACRO_MOUSE ACROSIN PRECURSOR (EC	1.21e-121
17	573	9.1	562	1 UROT_HUMAN TISSUE PLASMINOGEN ACT	8.37e-121
18	551	8.7	433	1 UROK_PAPCY UROKINASE-TYPE PLASMIN	1.21e-114
19	543	8.6	394	1 URTG_DESRO SALIVARY PLASMINOGEN A	2.07e-112
20	545	8.6	790	1 PLMN_PIG PLASMINOGEN (EC 3.4.21	5.72e-113
21	539	8.5	421	1 ACRO_HUMAN ACROSIN PRECURSOR (EC	2.70e-111
22	530	8.4	333	1 PLMN_CANFA PLASMINOGEN (EC 3.4.21	8.68e-109
23	529	8.4	442	1 UROK_PIG UROKINASE-TYPE PLASMIN	1.65e-108

RESULT	1	STANDARD:	PRT:	875 AA.	ALIGNMENTS
ID	NETR_HUMAN				
AC	P56730:				
DT	15-DEC-1999 (rel. 39, Created)				
DT	15-DEC-1999 (rel. 39, Last sequence update)				
DT	15-DEC-1999 (rel. 39, Last annotation update)				
DE	NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).				
GN	PRSS12.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Homidae; Homo.				
NC	[1]				
RP	SEQUENCE FROM N.A.				
RP	TISSUE-BRAIN:				
RX	MEDLINE: 98201705.				
RA	PROBA K., GSCHWEND T.P., SONDERGGER P.;				
RT	"Cloning and sequencing of the cDNA encoding human neutrotypin.";				
RL	Biochim. Biophys. Acta 1396:143-147(1998).				
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC				
CC	ACTION MAY SUBSEQUE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH				
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.				
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/				
CC	or send an email to license@isb-sdb.ch).				
CC	EMBL: AJ001531; CAA04816.1; -				
DR	PROSITE: PS00134; TRYPIN_HIS: 1.				
DR	PROSITE: PS00135; TRYPIN_SER: 1.				
DR	PROSITE: PS00420; SPERACT_RECEPTOR: 3.				
KW	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.				
FT	SIGNAL	1	20	POTENTIAL.	
FT	CHAIN	21	875	NEUTROTYPIN.	
FT	DOMAIN	23	92	PROLINE-RICH.	
FT	DOMAIN	93	165	KRINGLE.	
FT	DOMAIN	170	271	SRCR 1.	
FT	DOMAIN	280	381	SRCR 2.	
FT	DOMAIN	387	487	SRCR 3.	

50	SEQUENCE	761 AA:	84118 MW;	3FC4F35 CRC32;
	Query Match	66.4%;	Score 4199;	DB 1; Length 761;
	Best Local Similarity	87.0%;	Pred. No. 0.00e+00;	
	Matches	528; Conservative	51; Mismatches	28; Indels 0; Gaps 0;
Dd	155	CDCGGGALPYRLVGVNGSGHEGRVELYHAGQWITICDQMNNADADVLCROGLSLGIK	214	
Oy	216	CSFSHGPFPIIRLLAGSSVHGEGRVELYHAGQWITICDQMNDADAELYCRQLGSLGIK	275	
Dd	215	AHWQHAFEGSGPILDEVRCTGNELSTIQCKSSWGHNCHKEKDAGVSCPLTDGVIR	274	
Oy	276	AMHQAYFEGBSPVMALDEVKTCTGNELSTIQCKSSWGHNCHKEKDAGVSCPLTDGVIR	335	
Dd	275	IAGKSTHEGRLEVYYKGQMGTVCDGWTEMTTYACRLGLGRKYGOSSVNHFDSNRP1	334	
Oy	336	IAGGSGSHGGRLEVYYRGQMGTVCDGWTELTNTYYVCRQLGFRKYGOASANHFEBSTGP1	395	
Dd	335	WLDDVSCKGKEVSIQCSTRQWRDRDCSHREDVGLTCYDSOGHLRSPFPRLVDGENK	394	
Oy	396	WLDDVSCKGKEVSIQCSTRQWRDRDCSHREDVSIACYPGGGHRLSLGFVRLMDGENK	455	
Dd	395	KEGRVEEVNQGMGTICDDGMTDKAAAVICROLGYKKPARATMAVFGEKGPIMHDNYK	454	
Oy	456	KEGREVEIINGOMGATICDDGMTDKAAVICROLGYKKPARATMAVFGEKGPIMHDNYK	515	
Dd	455	CTGENKALADVCYKODIGRHNRHSBDAVICYDLEKKASSGNKEMLSSGCGLRLHRRO	514	
Oy	516	CTGNRSRLADCIKODIGHNRHSBSDAVICYDFGKASKGNKESLSVCGCLRLLHRRO	575	
Dd	515	KRIIGGNSLGRAMWQASLRSAHGDRLCGATLLSSCWVLTAHCFKRYGNSREY	574	
Oy	576	KRIIGKNSSLRGWQWQVSLRKSSHGDRLCGATLLSSCWVLTAHCFKRYGNSREY	635	
Dd	575	AVRVADYTLVPEEFEEQEIVQOVIYIHRRYPDRSDYDALVRLQOPGCCARLSTHVLP	634	
Oy	636	AVRVADYHTLVPEEPEEELGVQOVIYIHRYPDRSDYDALVRLQOPPEEQCARFSSHVP	695	
Dd	635	ACLPLMRERPOKTASNCHITTGMDTGRAYSRTLOOAAPVLLPKRFCKERYKGLFTGRMLC	694	
Oy	696	ACLPLMRERPOKTASNCHITTGMDTGRAYSRTLOOAAPVLLPKRFCKERYKGLFTGRMLC	755	
Dd	695	AGNLQEDNRNVDSGGDSGGPLMCEPRDESWWYVYTSMCYGCGVKDTPEGVTVPAPVFW	754	
Oy	756	AGNLHEHRKRNVDSGGDSGGPLMCEPRGESWVVYVYTSMWYGCGVXDPGVYTKVASAFVW	815	
	755	IKSVSTL 761		
	816	IKSVTKL 822		
RESULT	3	STANDARD:	PRT:	532 AA.
ID	SPEC	SRPU		
AC	p16264:			
Dt	01-AUG-1990	(Rel. 15, Created)		
Dt	01-AUG-1990	(Rel. 15, Last sequence update)		
Dt	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	EGG PEPTIDE SPEPACT RECEPTOR PRECURSOR.			
OS	Strongylocentrotus purpuratus (Purple sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;			
CC	Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;			
CC	Strongylocentrotus.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.			
RX	MEDLINE: 89184581.			
RA	DANGOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;			
RT	"Cloning of the mRNA for the protein that crosslinks to the egg			
RL	peptide speract."			
CC	Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).			
CC	-1- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPEPACT.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright It is produced through a collaboration			

Query	Match	Similarity	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231	Length	DB 232	Score	DB 233	Length	DB 234	Score	DB 235	Length	DB 236	Score	DB 237	Length	DB 238	Score	DB 239	Length	DB 240	Score	DB 241	Length	DB 242	Score	DB 243	Length	DB 244	Score	DB 245	Length	DB 246	Score	DB 247	Length	DB 248	Score	DB 249	Length	DB 250	Score	DB 251	Length	DB 252	Score	DB 253	Length	DB 254	Score	DB 255	Length	DB 256	Score	DB 257	Length	DB 258	Score	DB 259	Length	DB 260	Score	DB 261
-------	-------	------------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------

FT	DISULFID	92	109	BY SIMILARITY.
FT	DISULFID	111	120	BY SIMILARITY.
FT	DISULFID	128	209	BY SIMILARITY.
FT	DISULFID	149	191	BY SIMILARITY.
FT	DISULFID	180	204	BY SIMILARITY.
FT	DISULFID	219	300	BY SIMILARITY.
FT	DISULFID	240	282	BY SIMILARITY.
FT	DISULFID	271	295	BY SIMILARITY.
FT	DISULFID	303	434	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	346	362	BY SIMILARITY.
FT	DISULFID	354	423	BY SIMILARITY.
FT	DISULFID	448	523	BY SIMILARITY.
FT	DISULFID	480	496	BY SIMILARITY.
FT	DISULFID	513	541	BY SIMILARITY.
FT	CARBOHYD	153	153	POTENTIAL.
FT	CARBOHYD	487	487	POTENTIAL.
SEQ	SEQUENCE	566 AA;	63701 MW;	59A1E93B CRC32;

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2.30e-130;	Mismatches 107;	Indels 22; Gaps 17

QY	Local Similarity	9.6%	Score 607;	DB 1;	Length 566;
QY	Conservative	116;	Pred. NO. 2		

```

CC      -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC      THE MOLECULE INTO A LIGHT CHAIN WHICH CONTAINS THE ACTIVE SITE,
CC      AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KINOGEN. THESE
CC      CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC      -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPsin FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaborat
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commerc
CC      entities requires a license agreement (See http://www.isb-sib.ch/announ
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL; M58588; AAA63393.1;
CC      DR      PIR; A36557; KOMSPL.
CC      DR      HSSP; P00750; IRRF.
CC      DR      MGD; MGI:102849; KLK3.
CC      DR      PROSITE; PS00134; TRYPsin.HIS. 1.
CC      DR      PROSITE; PS00135; TRYPsin.SER. 1.
CC      DR      PROSITE; PS00495; APPLE. 4.
CC      DR      PFAM; PF00024; apple. 4.
CC      DR      PFAM; PF00089; trypsin. 1.
CC      DR      HydroLase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
CC      KW      Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
CC      KW      Duplication; Bradykinin.
CC      FT      SIGNAL          1          19
CC      FT      CHAIN          20          390
CC      FT      CHAIN          391          638
CC      FT      REPEAT          20          105
CC      FT      REPEAT          110          195
CC      FT      REPEAT          200          285
CC      FT      REPEAT          291          376
CC      FT      DOMAIN          389          621
CC      FT      CARBOHYD        127          127
CC      FT      CARBOHYD        308          308
CC      FT      CARBOHYD        396          396
CC      FT      CARBOHYD        453          453
CC      FT      CARBOHYD        494          494
CC      FT      ACCT_SITE        434          434
CC      FT      ACCT_SITE        483          483
CC      FT      ACCT_SITE        578          578
CC      FT      DISULFID        21          104
CC      FT      DISULFID        47          77
CC      FT      DISULFID        51          57
CC      FT      DISULFID        111          194
CC      FT      DISULFID        137          166
CC      FT      DISULFID        141          147
CC      FT      DISULFID        201          284
CC      FT      DISULFID        227          256
CC      FT      DISULFID        231          237
CC      FT      DISULFID        292          375
CC      FT      DISULFID        318          347
CC      FT      DISULFID        322          328
CC      FT      DISULFID        340          345
CC      FT      DISULFID        383          503
CC      FT      DISULFID        419          435
CC      FT      DISULFID        517          584
CC      FT      DISULFID        548          563
CC      FT      DISULFID        574          602
CC      FT      DISULFID        638          71368
CC      FT      DISULFID        71368          8370CD2E CRC32;
CC      SQ      SEQUENCE
CC
CC      Query Match          9.6%; Score 609; DB 1; Length 638;
CC      Best Local Similarity 40.8%; Pred. No. 6.28e-131;
CC      Matches 100; Conservative 54; Mismatches 73; Indels 18; Gaps 16.
CC
CC      Db      390 RIVGCTNASLEEMPOVSLQYKLV-SQTH-CGSGIIGROWVLAHAQFQIPYDWMRI 447
CC      ||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC      577 RIIIGKNSLRGGMPOVSLRKSSHGDRLLCGATLLSSCWVLAHAQCFK--RYGNSTRS 634
CC
CC      Db      448 YGGLIS-V-SEIKTEPSSR-IKEIILHQEKVSEGNDAIILIKLQPLN-YTEF--QK- 500

```



```

QY 635 YAVRVDYHTLVPEEPEEEIGVOQIVIHREYRPDRSDYDIALVLOGPEEQCARFSSHVL 694
Db 501 PICLP-SKADINTIYNWVTGNGYTKEOGETONILQKATIPLVNPECQKKYRDYVINK 559
QY 695 PACLPIMRERPOKTASNCYITGMDT-GRAYSR-LOQAIPILPKRCEERYKGRFTGR 752
Db 560 OMICAGT-KEGG-TDACKGDSGGPLVCKHSG-RWOLVGITSMGEGCGRKDPGVYTKVSE 616
QY 753 -MLCAGNLHEHKKRRVDSOGDSGGLMCEBPESWVYGYTSMGCGGVKDSPGVYTKVSA 811
Db 617 YMDMI 621
QY 812 FVPWI 816

```

h completed: Mon Mar 13 10:05:34 2000
ime : 37 secs.

THIS PAGE BLANK (USPTO)


```

      264 ICQDLGSLGAKAMHQAIFBEGSPVLMDEVRKGNELISTQCKRSSGHEINCGHKEDAG 323
      669 VVCASITSGFVGYTLRVGSGNSRQRLLEISINNMGTVCDDSDMDATVATVCRQLGFS 728
      324 VSCGPLDVG-VI-----RLAGKSGHEGRLEVEYHVGWGTVCDDMTLNTLVVCRQLGFK 378
      729 SAVSAPLSAHFGGSGITWLDVSCAGNENSLMDCGHRGLGVHNCANAEADAGVCIAS-D 787
      379 YG-KOASANHFEEESTGPIWLDVSCSKETRFLOCRRMWRHDCSHREDVSIACYPGE 437
      788 G---PLN--IRLGGREGMEGRVEISLGGDWGTVCDDSWGIEDAHVVCROLGF-CPALSA 841
      438 GHLSLGFPRLMDGKKEKREVEVFINGOMGTICDDGWTMDKDAVICROLGYGPA-RA 496
      842 VTAASFQGGSGILMDVQCSGDEATIAESCHNGIGHNCQHOEDAGVVC 891
      497 RTMAVFGEGKPIHVDNVKCTGNERSLADCIKODIGRHNCRHSDAGVYC 546

      2
      096943 PRELIMINARY: PRT; 2043 AA.
      096943:
      AC 01-MAY-1999 (TREMBLrel. 10, Created)
      DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
      DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
      DE SRCR DOMAIN, MEMBRANE FORM 2.
      GN SRCR2.
      OS Geodia cydonium (Sponge).
      OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
      RN Astrophorida; Geodiidae; Geodia.
      RN [1]
      RP SEQUENCE FROM N.A.
      RX MEDLINE: 96369060.
      RA BLUMBACH B., PANCER Z., DIEHL-SEIFERT B., STEFFEN R., MUENKNER J.,
      RA MUELLER I., MUELLER W.E.G.;
      RT "The putative sponge aggregation receptor. Isolation and
      RT characterization of a molecule composed of scavenger receptor
      RT cysteine-rich domains and short consensus repeats."
      RL J. Cell Sci. 111:2635-2644 (1998).
      DR EMBL: Y14953; CAAT5175.1;
      DR HSSP: P10998; IVCV.
      DR PROSITE: PS00420; SPERACT RECEPTOR: 8
      SQ SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;

```

Query Match 20.3%; Score 1285; DB 5; Length 2043;
 Best Local Similarity 43.4%; Pred. No. 2,87e-299;
 Matches 192; Conservative 86; Mismatches 150; Indels 14; Gaps 13;

```

      823 PIRLVNNGVITONEGVELLHNGTMSAVCDYWGYTEAVVACHMLGFATVRAVYTSF 882
      116 SVRL-RGGR--NEFEETVEYASGVWGTVCSSHMDSDASVICHQDLGGKIQKQTF 172
      883 GAVDGTDFLDNVACAGEREIFDCYHSSYISRNCDRTQOAGVACTNTATSEPI-RLVGG 941
      173 GGLRLTYNSVNCRGDEEWILLCEKDIWOGVCPQKMAAAVTSFSGHPPIIRLAGG 232
      942 GSPHEGRVEIYGVWGTVCDDSWGQPDADVCRQLGYANASRAVRAEFGRTGEIMLD 1001
      233 SSVHEBRVELYHAGWGTVCCDDQMDADAEVICROLGLSIGAKAMHQAIFBEGSPV 292
      1002 NVACTGFENSLDECRSNGWGDHNCGRHEDAGAVCQGGELL-IRLRDGSNLEGEVEYFN 1060
      293 EVRCTGENELIEQCPKSSWGEHNCGRHEDAGVSCPTPLTDVIRLAGKSGHGRLEVEY 352
      1061 KTMGTVCDDFW-DLGATVYCRQLGYPEAVRAEPARFAGSGEGRIIMDDVACVTEFS 1119
      353 GGMGTVCDDGWEL-NYYVCRQLGFEYKQASA-NHFESTGPIWLDVSCSKETRF 410
      1120 ECSSSGLGQHNCHQFEDAGVGN-DHITPS-PVT-DLRLAGTATGATGVEIFPNNMTGT 1176
      411 QCSRRMGHNDGSHREDVSIACIPGGGHRSLSLGFPRLMDGKKEKREVEVFINGOMGT 470

```

```

      1177 ICDSWIRDAEYVYCRGLGESAIEALSNGYFGAGDPDPITWLDVDCRESEITITSC 1236
      471 ICDDGWTMDKDAAVICROLGYKGPARTMAVFGEGK-G-PIHVDNVKCTGNERSLADCI 528
      1237 SALGHNCAHYEDAGVRC-YI 1257
      529 QDIRHNCRHSDAGVICYDFG 550

```

```

      3
      097378 PRELIMINARY: PRT; 1036 AA.
      097378:
      AC 01-MAY-1999 (TREMBLrel. 10, Created)
      DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
      DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
      DE SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN PRECURSOR.
      OS Strongylocentrotus purpuratus (Purple sea urchin).
      OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
      OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
      RN Strongylocentrotus.
      RN [1]
      RP SEQUENCE FROM N.A.
      RA PANCER Z., RAST J.P., DAVIDSON E.H.;
      RT "Origins of Immunity: Transcription Factors and Effector Genes of the
      RT Vertebrate Immune System Expressed in Sea Urchin Coelomocytes."
      RL Immunogenetics 0:0-0(1999).
      DR EMBL: AF076513; AAD05493.1;
      DR HSSP: P56682; IVCV.
      KM Signal; Receptor.
      FT SIGNAL 1 20
      FT CHAIN 21 1036
      FT SEQUENCE 1036 AA; 109704 MW; CB8C531 CRC32;

```

Query Match 19.2%; Score 1216; DB 5; Length 1036;
 Best Local Similarity 45.3%; Pred. No. 1,15e-280;
 Matches 199; Conservative 84; Mismatches 134; Indels 22; Gaps 14;

```

      615 QC-QGAIRLVGSNEAGREIY-NGWGTICDDSWGITDASVCRMFGOGASGAPGS 672
      111 DCRGSRVRLGKNEFEIYEV-YASGVWGTVCSSHMDSDASVICHQDL-GGKIQAKQ 168
      673 AHF-GGCTGPIQDDVCGTGAETFDCAHAFGVNCAHYEDAGVFC-IA-SQD---VR 726
      169 TPTSGDLPIYMSVNCRGDEEWILLCEKDIWOGVCPQKMAAAVTSFSGHPPIIR 228
      727 LVGSNEAGREIYGVWGTVCDDSWGITDASVCRMFGOGASGAPGSQFGGTGL 786
      229 LAGSSVHEBRVELYHAGWGTVCCDDQMDADAEVICROLGLSIGAKAMHQAIFBEGSGP 288
      787 IQDDVCGTGAETFDCAHAFGVNCAHYEDAGVCTA-SQDV-RLVGSNEAGREY 844
      289 VMLDEVCTGENELIEQCPKSSWGEHNCGRHEDAGVSCPTPLTDVIRLAGKSGHGRLE 348
      845 IQYVWGTICDDSWGITDANVCMIGFQASGAGSAGFQGTGPIQDDVCGTGVQ 904
      349 VYIRGQGTVCDDGWELNYYVCRQLGFEYKQASAN-HFEESTGPIWLDVSCSGKGT 407
      905 TIEPCAPPFVHNCAHYEDAGVCIYS---Q---D--VRLVDGNSMAAEGREIYNGV 955
      408 RFLQCSRRMGHNDGSHREDVSIACIPGGGHRSLSLGFPRLMDGKKEKREVEVFINGO 467
      956 WGTICDDFWITDANVCRMFGOGASGAGSAGFQGTGPIQDDVCGTGAETFDCA 1015
      468 WGTICDDGWITDKDAAVICROLGYKGPARTMAVFGEGKPIHVDNVKCTGNERSLADCI 527
      1016 HPPPGVNCRHSDAGVVC 1034
      528 KODIGRHNCRHSDAGVYC 546

      4
      007898 PRELIMINARY: PRT; 1116 AA.
      007898:

```

[illegible]

DB	392	VCBRCGMLKEDVYVCRQLGCGSAAKTSQVYSKIOATFTWFLSSCGNENSLMDCKNQ	451
QY	471	ICDDGWDKDAVTCRQLGKGPARTMAVFGEGKPIHVDNVCYGTGENSLADCIKOD	530
Db	452	WGGLTCDHYEBAKTC	467
QY	531	IGRNHCRHSEDAGYIC	546
RESULT	5	PRELIMINARY;	PRT: 1149 AA.
ID	Q07901		
AC	Q07901:		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)		
DE	M130 ANTIGEN, EXTRACELLULAR VARIANT PRECURSOR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 93380506.		
RA	LAW S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,		
RA	MAISON D.Y.:		
RT	"A new macrophage differentiation antigen which is a member of the		
RT	scavenger receptor superfamily."		
RL	Eur. J. Immunol. 23:2320-2325(1993).		
DR	EMBL: Z22971; CAAB0544.1; -		
DR	PFAM: PF00530; SRCR: 9.		
DR	PRINTS: PR00258; SPBRACRCPTR.		
KM	Antigen: Signal.		
FT	SIGNAL	40	POTENTIAL.
FT	CHAIN	41 1149	M130 ANTIGEN, EXTRACELLULAR VARIANT.
SEQ	SEQUENCE	1149 AA; 124328 MW; 4901C708 CRC32;	
Query Match		14.9%; Score 946; DB 4; Length 1149;	
Best Local Similarity		39.2%; Pred. No. 2,40e-208;	
Matches	171; Conservative	73; Mismatches 172; Indels 20; Gaps 16;	
Db	46	LVLVGKRCGAREVKKQOE ¹ GTVCNNNGMEAVSVCNOLG-CPTAIAFGWANSAG	104
QY	117	VLRLGKRFEGGYEVYASGWSGYVCSHWDSDASVICHOLDLGGKIAKQ ² -PESGLG	175
Db	105	SGRI-WMDVSCRGESALMDCKHDGKHSNCTHODAGVTS--DGSNLEM-RLTRGG	160
QY	176	LPIYW-SNVRCRGDEEILLCEKDIW-QGGVCPQKMAAVTCSFSGPPEPIRLAGGS	233
Db	161	NMCSRIIRIKQ-GRMGVCDNDNFNIDHASYICQLE-CGSNVSFSSSNFGESGPIWF	218
QY	234	SVHERVEL-YHAQMGVTCDDDDADAELVICHOLDLGAKMWH-QAYFGEBSGVM	291
Db	219	DDLJNGESALMNCKHQMGKHNCDHAEDAGVYCSGADLSLRLVGYECSGRLEVRF	278
QY	292	DEVRCITGELSIEDCPSSMGEHNCGKEDAGVSCPTLTGVIRLAGKSSHEGRLEVTY	351
Db	279	QEGMGITICDDGWSYDAVAVKOLGCEPTAVATAGRVNASKGFHIMWDSVSCGHEPAW	338
QY	352	RGVNGVTDGDMTELTNYVVCRLGFRFYKQASNN-HFESTGPIWLDVSCSGKETRL	410
Db	339	QCKHHEGKATCNHEDAGVTC-SDG-SD-LEL---RLRGGSRCAGIVEIQRLLGK	391
QY	411	QCSRWOMGRHDCSHREVYSIACYGEGGHRILSLGFPYRLMDGEEKKGEVFIINGOMGT	470
Db	392	VCDGWMGLKEDVYVCRQLGGSAAKTSQVYSKIOANTWFLSSCGNENSLMDCKNQ	451
QY	471	ICDDGWDKDAVTCRQLGKGPARTMAVFGEGKPIHVDNVCYGTGENSLADCIKOD	530
Db	452	WGGLTCDHYEBAKTC	467
QY	531	IGRNHCRHSEDAGYIC	546
RESULT	6		

```

ID 007899 PRELIMINARY: PRT: 1151 AA.
AC Q07899;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAW S.A., MICKLEW K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22969; CAAB0542.1; -.
DR PFM: PF00530; SRCR: 9.
DR PRINTS: PRO0258; SPERACTRCPTR.
FI SIGNAL: 1 40 POTENTIAL.
SI CHAIN 1 1151 M130 ANTIGEN, CYTOPLASMIC VARIANT 1.
SQ SEQUENCE 1151 AA; 124820 MW; A72EDD2F CRC32;

Query Match 14.9%; Score 946; DB 4; Length 1151;
Best Local Similarity 39.2%; Pred. No. 2,406-208;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVGENKSGRVEYKVOEEMGTVCNNGSMSEAVSICNOLG-CPTAIKAPGMANSSAG 104
117 VRLRGKNEFEETVEYASGVWGTVCSSHMDDSDASYICHOLOGKGIKOT-PFSGIG 175
105 SGRV-WMDHVSGRGNSALMDCKHDGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
176 LIPIY-SVNRGRGDEENILCEKDIW-QGVCPQKMAAAVTCFSHGPTPIIRLAGGS 233
161 NMCSRIEIKFO-GRMGVTCDDNFNIDHASYICROLE-CGSNVSFGSSNFGSGSPITF 218
224 SVHEGRVEL-YHAGQGTVCDDQWMDADAEVICRQLGSLGIAKAMH-QAYFGSGSPVML 291
219 DDLICNGNSALMNCKHOGKHNCDHAEDAGVTCGADLSRLVDGVTESGRLEVR 278
232 DEVRTGNELEIEQCPKSSWGEHNCHEKEDAGVCTPLTDGVIKRLGGSGHEGRLEVIY 351
279 QGEMGTICDDGSDSYDAVAACQOLGCPYAVTAIGRVNASKGFEGHILWDSYSCQHEPVAW 338
352 RGQWGTVCDDGWTETLNTYVVCROLGFKYKQASAN-HFEESTGPIWLDVDSGSGKTRFL 410
339 QCKHHEMGKHYCNHNEADAGVTC-SDG-SP-LEL---RLRGGSRCAGTVEYETORLLGK 391
411 QCSRQWGRHDCSHREDYSIACYPGEGHRLSLGFPVRLMDGKREKREGEVYFINGQWGT 470
392 VCDRGMGLKEADVVCROLGCGSALKTSYQVYSKIQTNTWFLSSCNGNETSLMDCKNMQ 451
471 ICDDGWTDKDAVVICROLGKQARARTMAVFEKGKPIHVDNVKCTGNERSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRHNCRHSDEAGVTC 546

```

```

RESULT 7
ID 007900 PRELIMINARY: PRT: 1156 AA.
AC Q07900;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA LAW S.A., MICKLEW K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily.";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: Z22970; CAAB0543.1; -.
DR PFM: PF00530; SRCR: 9.
DR PRINTS: PRO0258; SPERACTRCPTR.
DR Antigen: Signal.
FI SIGNAL: 1 40 POTENTIAL.
SI CHAIN 1 1156 M130 ANTIGEN, CYTOPLASMIC VARIANT 2.
SQ SEQUENCE 1156 AA; 125352 MW; 287A07A0 CRC32;

Query Match 14.9%; Score 946; DB 4; Length 1156;
Best Local Similarity 39.2%; Pred. No. 2,406-208;
Matches 171; Conservative 73; Mismatches 172; Indels 20; Gaps 16;

Db 46 LRLVGENKSGRVEYKVOEEMGTVCNNGSMSEAVSICNOLG-CPTAIKAPGMANSSAG 104
117 VRLRGKNEFEETVEYASGVWGTVCSSHMDDSDASYICHOLOGKGIKOT-PFSGIG 175
105 SGRV-WMDHVSGRGNSALMDCKHDGKHSNCTHOODAGVTC--DGSNLEM-RLTRGG 160
176 LIPIY-SVNRGRGDEENILCEKDIW-QGVCPQKMAAAVTCFSHGPTPIIRLAGGS 233
161 NMCSRIEIKFO-GRMGVTCDDNFNIDHASYICROLE-CGSNVSFGSSNFGSGSPITF 218
224 SVHEGRVEL-YHAGQGTVCDDQWMDADAEVICRQLGSLGIAKAMH-QAYFGSGSPVML 291
219 DDLICNGNSALMNCKHOGKHNCDHAEDAGVTCGADLSRLVDGVTESGRLEVR 278
292 DEVRTGNELEIEQCPKSSWGEHNCHEKEDAGVCTPLTDGVIKRLGGSGHEGRLEVIY 351
279 QGEMGTICDDGSDSYDAVAACQOLGCPYAVTAIGRVNASKGFEGHILWDSYSCQHEPVAW 338
352 RGQWGTVCDDGWTETLNTYVVCROLGFKYKQASAN-HFEESTGPIWLDVDSGSGKTRFL 410
339 QCKHHEMGKHYCNHNEADAGVTC-SDG-SP-LEL---RLRGGSRCAGTVEYETORLLGK 391
411 QCSRQWGRHDCSHREDYSIACYPGEGHRLSLGFPVRLMDGKREKREGEVYFINGQWGT 470
392 VCDRGMGLKEADVVCROLGCGSALKTSYQVYSKIQTNTWFLSSCNGNETSLMDCKNMQ 451
471 ICDDGWTDKDAVVICROLGKQARARTMAVFEKGKPIHVDNVKCTGNERSLADCIKOD 530
Db 452 WGLTCDHYEAKITC 467
QY 531 IGRHNCRHSDEAGVTC 546

```

Search completed: Mon Mar 13 10:07:18 2000
Job time : 86 secs.

 W30E34 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 _pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:22:30 2000; Maspar time 5.73 Seconds
 Tabular output not generated. 301.681 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (40-112) From US09147947A.ppt (2 of 6)
 Perfect Score: 611
 Sequence: 1 CPAGEPWSVTDFGAPCLRM.....RPMCFYGDARKVDMGYCDC 73

Scoring table: PAM 150
 Gap 11
 Searched: 188963 seqs, 23686106 residues
 Post-processing: Minimum Match 04
 Listing first 45 summaries

Database: a-geneseq35
 i-geneseqp

Statistics: Mean 27.866; Variance 105.747; scale 0.264

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	611	100.0	822	1 W99087 Human serine protease	1.69e-54
2	611	100.0	875	1 W83361 Mouse neurotrypsin.	1.69e-54
3	507	83.0	761	1 W83362 Mouse serine protease	4.16e-43
4	307	83.0	761	1 W99088 Tissue plasminogen act	4.16e-43
5	319	31.9	439	1 W52815 Tissue plasminogen act	5.34e-10
6	319	31.9	439	1 W52814 Tissue plasminogen act	5.34e-10
7	319	31.9	527	1 W54151 Tissue plasminogen act	5.34e-10
8	319	31.9	527	1 W52813 Tissue plasminogen act	5.34e-10
9	319	31.9	527	1 W54150 t-PA mutant (S127C, Y1	5.34e-10
10	319	31.9	527	1 W54158 t-PA mutant (K159R).	5.34e-10
11	319	31.9	527	1 W52817 Tissue plasminogen act	5.34e-10
12	319	31.9	527	1 W54157 t-PA mutant (N142S).	5.34e-10
13	319	31.9	527	1 W54156 t-PA mutant (N115S).	5.34e-10
14	319	31.9	527	1 W54148 t-PA mutant (N115S, G1	5.34e-10
15	319	31.9	527	1 W54155 t-PA mutant (Y156D).	5.34e-10
16	319	31.9	527	1 W54154 t-PA mutant (N142S).	5.34e-10
17	319	31.9	527	1 W54153 t-PA mutant (G137D, W1	5.34e-10
18	319	31.9	527	1 W54152 t-PA mutant (R130Q, K1	5.34e-10
19	319	31.9	527	1 W54149 t-PA mutant (N146I, G1	2.69e-09
20	30.8	30.8	527	1 W54147 t-PA mutant (Y93C, T10	2.69e-09
21	187	30.6	483	1 R70845 Human tissue PA varian	3.39e-09
22	187	30.6	483	1 R70880 Human tissue PA varian	3.39e-09
23	187	30.6	483	1 R70887 Human tissue PA varian	3.39e-09

24 187 30.6 483 1 R09250 t-PA variant d1-44, V2 3.39e-09
 25 187 30.6 483 1 R70852 Human tissue PA varian 3.39e-09
 26 186 30.4 439 1 R09259 t-PA variant d92-179, 4.27e-09
 27 186 30.4 483 1 R09254 t-PA variant d1-44, I2 4.27e-09
 28 186 30.4 483 1 R70853 Human tissue PA varian 4.27e-09
 29 186 30.4 483 1 R70881 Human tissue PA varian 4.27e-09
 30 186 30.4 483 1 R70888 Human tissue PA varian 4.27e-09
 31 186 30.4 483 1 R09248 t-PA variant d1-44, I2 4.27e-09
 32 186 30.4 483 1 R70847 Human tissue PA varian 4.27e-09
 33 186 30.4 527 1 R09255 t-PA variant I210R, G2 4.27e-09
 34 186 30.4 527 1 R09255 t-PA variant d1-44, V2 6.76e-09
 35 184 30.1 483 1 R09252 t-PA variant d92-179, 8.51e-09
 36 183 30.0 439 1 R09260 t-PA variant d92-179, 8.51e-09
 37 183 30.0 439 1 R09261 t-PA variant d1-44, N1 8.51e-09
 38 183 30.0 483 1 R09257 Thrombolytic protein w 8.51e-09
 39 183 30.0 527 1 R05806 Thrombolytic protein w 8.51e-09
 40 183 30.0 528 1 R07033 Thrombolytic protein w 8.51e-09
 41 183 30.0 530 1 P92277 Sequence of modified t 8.51e-09
 42 183 30.0 531 1 R07034 Thrombolytic protein w 8.51e-09
 43 183 30.0 562 1 R07079 Thrombolytic protein w 8.51e-09
 44 181 29.6 483 1 R09249 t-PA variant d1-44, I2 1.35e-08
 45 180 29.5 356 1 W28535 mt-PA6-D. 1.69e-08

ALIGNMENTS

RESULT 1
 ID W99087 standard; Protein: 822 AA.
 AC W99087:
 DT 13-MAY-1999 (first entry)
 DE Human serine protease BSSP-3.
 KW Serine protease; BSSP-3; brain tissue.
 OS Homo sapiens.
 PN W09905290-A1.
 PD 04-FEB-1999.
 PF 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SUNR) SUNTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 DR N-PSDB: X19024.
 PT New serine protease expressed in brain tissue - used in screening
 for potential serine protease inhibitors for drug use
 PS Claim 1; Page 61-65; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 is isolated from human brain tissue. Transformants may be used to
 produce the enzyme or its partial sequences. Products from the present
 invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 822 AA;

Query Match 100.0%; Score 611; DB 1; Length 822;
 Best Local Similarity 100.0%; Pred. No. 1.69e-54;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 CPAGEPWSVTDFGAPCLRMWVPEFLERSPPASNAOLRGORHNCSPDGPAPCFYGC 99
 40 CRAGEPWSVTDFGAPCLRMWVPEFLERSPPASNAOLRGORHNCSPDGPAPCFYGC 99
 QY 40 CRAGEPWSVTDFGAPCLRMWVPEFLERSPPASNAOLRGORHNCSPDGPAPCFYGC 99
 Db 100 DARGKVDWGYCDC 112
 QY 100 DARGKVDWGYCDC 112

RESULT 2
 ID W83361 standard; Protein: 875 AA.
 AC W83361:
 DT 17-FEB-1999 (first entry)
 DE Human neurotrypsin.
 KW Human; neurotrypsin; tumour inhibition; neurological disease;
 lung disease; gene therapy; drug development; stroke; brain injury;
 neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

W30E34
At Nla 5 may be available

KV emphysema; bronchitis.
 OS Homo sapiens.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotrophin - used, e.g. for inhibiting tumours, treatment of neurological or lung disease, including by gene therapy and in drug development.
 PS Claim 1: Page 20-24; 50pp; English.
 CC The present sequence represents human neurotrophin. Neurotrophin proteins and polynucleotides can be used: (i) to inhibit tumours, including metastases, e.g. of brain or retina; (ii) to minimise tissue damage caused by stroke or brain injury (having a protective effect on the penumbra zone); (iii) to treat or prevent neurodegeneration, neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to increase survival of damaged neurons (e.g. in cases of hypoxia, ischaemia, nerve transection) and to stimulate regeneration and/or restoration of synapses; (v) to treat or prevent retinal disorders (e.g. degeneration or neovascularisation); (vi) to prevent apoptosis (or other causes of cell death) in the nervous system; (vii) to regenerate brain and/or nervous tissue; (viii) to treat pain; (ix) to improve brain performance, including learning and memory; (x) to treat or prevent a wide range of psychiatric disorders; and (xi) to treat brain or lung injury associated with protease expression (specifically emphysema or bronchitis).
 CC Sequence 875 AA:

Query Match 100.0%; Score 611; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 1.69e-54;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 93 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQRHNCRSPDGRMPCFTG 152
 QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQRHNCRSPDGRMPCFTG 99
 DB 153 DARGKVDWGYCDC 165
 QY 100 DARGKVDWGYCDC 112

RESULT 3
 ID W83362 standard; Protein; 761 AA.
 AC W83362.
 DT 17-FEB-1999 (first entry)
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72590.
 PT New human and murine neurotrophin - used, e.g. for inhibiting tumours, treatment of neurological or lung disease, including by gene therapy and in drug development.
 PS Claim 1: Page 29-32; 50pp; English.
 CC The present sequence represents mouse neurotrophin. Neurotrophin proteins and polynucleotides can be used: (i) to inhibit tumours, including metastases, e.g. of brain or retina; (ii) to minimise tissue damage caused by stroke or brain injury (having a protective effect on the penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to increase survival of damaged neurons, (e.g. in cases of hypoxia, ischaemia, nerve transection) and to stimulate regeneration and/or restoration of synapses; (v) to treat or prevent retinal disorders (e.g. degeneration or neovascularisation); (vi) to prevent apoptosis (or other causes of cell death) in the nervous system; (vii) to regenerate brain and/or nervous tissue; (viii) to treat pain; (ix) to improve brain performance, including learning and memory; (x) to treat or prevent a wide range of psychiatric disorders; and (xi) to treat brain or lung injury associated with protease expression (specifically emphysema or bronchitis).
 CC Sequence 761 AA:

Query Match 83.0%; Score 507; DB 1; Length 761;
 Best Local Similarity 78.1%; Pred. No. 4.16e-43;
 Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

DB 85 CGAGESMGATNIGVPCILMDEVPFLERSPPASMAELRGQRHNCRSPDGRMPCFTG 144
 QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQRHNCRSPDGRMPCFTG 99
 DB 145 NAQKVDWGYCDC 157
 QY 100 DARGKVDWGYCDC 112

RESULT 4
 ID W99088 standard; Protein; 761 AA.
 AC W99088.
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 OS Mus sp.
 PN W09905290-A1.
 PD 04-FEB-1999.
 PE 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SDNR) SUNTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR N-PSDB: X19027.
 PT New serine protease expressed in brain tissue - used in screening for potential serine protease inhibitors for drug use.
 PS Example 1: Page 51-54; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which is isolated from mouse brain tissue. Transforms may be used to produce the enzyme or its partial sequences. Products from the present invention are used for screening for potential peptide or non-peptide serine protease inhibitors or expression regulators for use as drugs.
 CC Sequence 761 AA:

Query Match 83.0%; Score 507; DB 1; Length 761;
 Best Local Similarity 78.1%; Pred. No. 4.16e-43;
 Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

DB 85 CGAGESMGATNIGVPCILMDEVPFLERSPPASMAELRGQRHNCRSPDGRMPCFTG 144
 QY 40 CPAGEPWSVTFGAPCLMWEVPPFLERSPPASMAQLRGQRHNCRSPDGRMPCFTG 99
 DB 145 NAQKVDWGYCDC 157
 QY 100 DARGKVDWGYCDC 112

RESULT 5
 ID W52815 standard; Protein; 439 AA.
 AC W52815.
 DT 07-JUL-1998 (first entry)
 DE Tissue plasminogen activator variant 2.
 KW tPA: fibrin-stimulated; clot; treatment; vascular disease;
 KW fibrin deposition; adhesion formation.
 OS Synthetic.
 OS Homo sapiens.

[illegible]

Accession	Protein Name	Location/Qualifiers
DE	Tissue plasminogen activator variant 1.	
KW	tPA: fibrin-stimulated; clot treatment: vascular disease; fibrin deposition; adhesion formation.	
KM	Synthetic.	
OS	Homo sapiens.	
FT	Key	
FT	Disulfide_bond	7..18
FT	Disulfide_bond	12..31
FT	Disulfide_bond	31..40
FT	Disulfide_bond	48..129
FT	Disulfide_bond	69..111
FT	Disulfide_bond	100..124
FT	Disulfide_bond	136..217
FT	Modified_site	140
FT	/note= "N-glycosylation"	
FT	Disulfide_bond	157..199
FT	Disulfide_bond	188..212
FT	Disulfide_bond	220..351
FT	Cleavage_site	233
FT	Disulfide_bond	263..279
FT	Disulfide_bond	271..340
FT	Disulfide_bond	365..440
FT	Disulfide_bond	397..413
FT	Disulfide_bond	430..457
FT	Domain	7..47
FT	/note= "contains the growth factor domain"	
FT	Domain	48..129
FT	/note= "Kringler-1 domain"	
FT	Domain	136..217
FT	/note= "Kringler-2 domain"	
FT	Domain	220..483
FT	/note= "serine protease domain"	
PN	US5714145-A.	
PD	03-FEB-1998.	
PR	02-SEP-1988; 240856.	
PR	24-JUL-1989; US-383608.	
PR	02-SEP-1988; US-240856.	
PR	03-OCT-1991; US-770510.	
PR	06-JUL-1993; US-088451.	
PR	07-JAN-1994; US-179059.	
PR	14-APR-1995; US-422736.	
PR	29-MAR-1996; US-622891.	
PR	17-OCT-1996; US-733353.	
PA	(GERTH) GENENTECH INC.	
PI	Anderson S, Bennett WF, Botstein D, Higgins DL, Paoni NF, Zoller MJ.	
PI	WPI: 98-129803/12.	
PT	Treatment of vascular conditions or disease - using tissue plasminogen activator variant having amino acid substitutions in protease domain to increase fibrin specificity	
PT	Claim 4; Page 7. 31pp; English.	
PS	Variant5 W52814-W52817 containing two or more functional regions were created from tissue plasminogen activator (tPA)(W52813). The t-PA variants have a higher fibrin-stimulated activity than fibrinogen-stimulated activity so they will act preferentially at the site of a clot and not systemically. They can be used for treating vascular diseases and conditions or to prevent fibrin deposition or adhesion formation or reformation.	
CC	Note: This sequence is not given in the specification but was created from the wildtype t-PA sequence (W52813) disclosed by the inventors.	
CC	Sequence 483 AA;	
SO		
Query Match	31.9%; Score 195; DB 1; Length 483;	
Best Local Similarity	39.4%; Pred. No. 5,34e-10;	
Matches	26; Conservative 15; Mismatches 22; Indels 3; Gaps 3.	
Db	149 SLTESGASCLPWNMILIGKYTAQNPSSAALGIGKHNCRNPDGAKPMCHYLKNR-RL 207	
Oy	48 SVTDFGACRLWAEVPEFLER-SPPASMAQLRQG-RHNFCRSPDGAGRPMCFGDARGKV 105	
Db	208 TWEXCD 213	
Oy	106 DWGYCD 111	

Db 252 TWGYCD 257
QY 106 DWGYCD 111

RESULT 14
ID W54148 standard; protein: 527 AA.

AC W54148;
DE 20-JUL-1998 (first entry)
KW t-PA mutant (M15S, G139R)
KW Amino acid substitution; t-PA; vascular disorder; prevention;
OS fibrin deposition; adhesion formation.
OS Synthetic.

Key Location/Qualifiers
Domain 1..44
/note= "Finger domain"
Domain 45..91
/note= "Growth factor domain"
Domain 92..173
/note= "Kringlet-1 domain"
Domain 180..261
/note= "Kringlet-2 domain"
Domain 264..527
/note= "Serine protease domain"
FT misc_difference 115
/note= "N changed from wt to S in mutant"
FT misc_difference 139
/note= "G changed from wt to R in mutant"
FT misc_difference 139
/note= "G changed from wt to R in mutant"
T. US5736135-A.
PD 07-APR-1998.
PR 13-FEB-1995; 389615.
PR 11-JUL-1991; US-728456.
PR 26-JAN-1993; US-008940.
PR 01-APR-1994; US-221660.
PR 13-FEB-1995; US-389615.
PA (GERTH) GENENTECH INC.
PI Goeddel DV, Leung DW, Rice GC;
DR WPI: 98-239153/21.
PT Mutant tissue plasminogen activator proteins - useful for treating
PT vascular disorders, preventing tissue adhesion(s), etc.
PS Claim 6; Page -: 24pp; English.
CC Mutant tissue plasminogen activator proteins (W54147-W54158) are created
CC by single or multiple amino acid substitutions. Compositions containing
CC the t-PA variant are used for treating vascular disorders, for preventing
CC fibrin deposition or for preventing adhesion formation or reformation.
Note: This sequence is not given in the specification but was created
from the wild type by the indexer.
Sequence 527 AA;

Query Match 31.9%; Score 195; DB 1; Length 527;
Best Local Similarity 39.4%; Pred. No. 5.34e-10;
Matches 26; Conservative 15; Mismatches 22; Indels 3; Gaps 3;

Db 193 SLTEGASCLPWNMSMLIGKVTYTAONPSAOLGLGKHNCRPDGDAKPMCHYLKRR-RL 251
QY 48 SVTDEGACPLRAEVPFLEER-SPRPSWAQLRGQ-RHNFCRSPDAGRPWCYTGDAKGV 105
Db 252 TWGYCD 257
QY 106 DWGYCD 111

RESULT 15
ID W54155 standard; protein: 527 AA.
AC W54155;
DE 20-JUL-1998 (first entry)
KW t-PA mutant (Y156D)
KW Amino acid substitution; t-PA; vascular disorder; prevention;
OS fibrin deposition; adhesion formation.
OS Synthetic.
FT Key Location/Qualifiers
FT Domain 1..44

FT Domain /note= "Finger domain"
FT Domain 45..91
FT /note= "Growth factor domain"
FT Domain 92..173
FT /note= "Kringlet-1 domain"
FT Domain 180..261
FT /note= "Kringlet-2 domain"
FT Domain 264..527
FT /note= "Serine protease domain"
FT misc_difference 156
/note= "Y changed from wt to D in mutant"
FN US5736135-A.
PD 07-APR-1998.
PR 13-FEB-1995; 389615.
PR 11-JUL-1991; US-728456.
PR 26-JAN-1993; US-008940.
PR 01-APR-1994; US-221660.
PR 13-FEB-1995; US-389615.
PA (GERTH) GENENTECH INC.
PI Goeddel DV, Leung DW, Rice GC;
DR WPI: 98-239153/21.
PT Mutant tissue plasminogen activator proteins - useful for treating
PT vascular disorders, preventing tissue adhesion(s), etc.
PS Claim 6; Page -: 24pp; English.
CC Mutant tissue plasminogen activator proteins (W54147-W54158) are created
CC by single or multiple amino acid substitutions. Compositions containing
CC the t-PA variant are used for treating vascular disorders, for preventing
CC fibrin deposition or for preventing adhesion formation or reformation.
CC Note: This sequence is not given in the specification but was created
from the wild type by the indexer.
SQ Sequence 527 AA;

Query Match 31.9%; Score 195; DB 1; Length 527;
Best Local Similarity 39.4%; Pred. No. 5.34e-10;
Matches 26; Conservative 15; Mismatches 22; Indels 3; Gaps 3;
Db 193 SLTEGASCLPWNMSMLIGKVTYTAONPSAOLGLGKHNCRPDGDAKPMCHYLKRR-RL 251
QY 48 SVTDEGACPLRAEVPFLEER-SPRPSWAQLRGQ-RHNFCRSPDAGRPWCYTGDAKGV 105
Db 252 TWGYCD 257
QY 106 DWGYCD 111

Search completed: Mon Mar 13 10:22:38 2000
Job time : 8 secs.

THIS PAGE BLANK (USPTO)

40-112/Kurugle
-147-947-6-02.ra1
PCT publication revision on file

(77)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
h_pp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Mar 13 10:22:56 2000;      MasPar time 4.22 Seconds
Tabular output not generated.      224.268 Million cell updates/sec
```

```

Title: .>US-09-147-947-6
Description: (40-112) from US09147947A pep (2 of 6)
Perfect Score: 611
Sequence: 1 CPAGPFWSTVDGFGAPCIKRW.....RPMCFYGDANGKVDWGXCDC 73

```

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

```
Post-processing: Minimum Match 08
                  Listing first 45 summaries
```

```
Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1
```

Statistics: Mean 26.490; Variance 105.747; scale 0.251

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	207	33.9	946	3	PRT-US95-0	Sequence 13, Applicant	5.54e-111
2	186	30.4	527	4	552013-1	Patent No. 552013.	5.92e-08
3	180	29.5	356	1	US-08-421	Sequence 4, Applicant	2.22e-08
4	179	29.3	347	2	US-08-427	Sequence 1, Applicant	2.77e-08
5	179	29.3	354	2	US-08-811	Sequence 61, Applicant	2.77e-08
6	179	29.3	355	2	US-08-811	Sequence 59, Applicant	2.77e-08
7	179	29.3	355	1	US-08-137	Sequence 1, Applicant	2.77e-08
8	179	29.3	355	4	522355-1	Patent No. 522355.	2.77e-08
9	179	29.3	355	2	US-08-811	Sequence 47, Applicant	2.77e-08
10	179	29.3	355	1	US-08-217	Sequence 1, Applicant	2.77e-08
11	179	29.3	355	2	US-08-811	Sequence 53, Applicant	2.77e-08
12	179	29.3	355	1	US-08-217	Sequence 1, Applicant	2.77e-08
13	179	29.3	355	1	US-08-427	Sequence 1, Applicant	2.77e-08
14	179	29.3	355	1	US-08-427	Sequence 2, Applicant	2.77e-08
15	179	29.3	355	1	US-08-427	Sequence 6, Applicant	2.77e-08
16	179	29.3	355	2	US-08-811	Sequence 45, Applicant	2.77e-08
17	179	29.3	389	2	US-08-811	Sequence 65, Applicant	2.77e-08
18	179	29.3	389	2	US-08-811	Sequence 67, Applicant	2.77e-08
19	179	29.3	437	2	US-08-811	Sequence 49, Applicant	2.77e-08
20	179	29.3	437	2	US-08-811	Sequence 55, Applicant	2.77e-08
21	179	29.3	437	2	US-08-811	Sequence 51, Applicant	2.77e-08
22	179	29.3	437	2	US-08-811	Sequence 57, Applicant	2.77e-08
23	179	29.3	472	2	US-08-811	Sequence 63, Applicant	2.77e-08

24	179	29.3	527	1	5185255-8	Patent No. 5185255.	2.77e-08
25	179	29.3	527	1	US-07-609-9	Sequence 16, Applicant	2.77e-08
26	179	29.3	527	3	PCT-US91-0	Sequence 2, Applicant	2.77e-08
27	179	29.3	527	2	US-08-811-6	Sequence 39, Applicant	2.77e-08
28	179	29.3	546	4	5200340-6	Patent No. 5200340.	2.77e-08
29	179	29.3	562	4	5185255-9-3	Patent No. 5185255.	2.77e-08
30	179	29.3	562	4	5200340-0-2	Patent No. 5200340.	2.77e-08
31	179	29.3	562	4	5246756-5	Patent No. 5246756.	2.77e-08
32	179	29.3	562	4	5344773-2	Patent No. 5344773.	2.77e-08
33	179	29.3	562	2	US-08-811-6	Sequence 43, Applicant	2.77e-08
34	179	29.3	562	2	US-08-883-0	Sequence 38, Applicant	2.77e-08
35	179	29.3	562	2	US-08-560-5	Sequence 50, Applicant	2.77e-08
36	177	29.0	78	2	US-08-612-2	Sequence 21, Applicant	4.30e-08
37	177	29.0	250	2	US-08-612-2	Sequence 33, Applicant	4.30e-08
38	177	29.0	339	1	US-08-326-6	Sequence 6, Applicant	4.30e-08
39	177	29.0	339	1	US-08-432-2	Sequence 6, Applicant	4.30e-08
40	177	29.0	339	1	US-08-248-8	Sequence 6, Applicant	4.30e-08
41	177	29.0	339	2	US-08-612-2	Sequence 6, Applicant	4.30e-08
42	177	29.0	339	2	US-08-866-6	Sequence 6, Applicant	4.30e-08
43	177	29.0	339	3	PCT-US95-0	Sequence 6, Applicant	4.30e-08
44	177	29.0	339	2	US-08-429-9	Sequence 6, Applicant	4.30e-08
45	177	29.0	339	1	US-08-451-1	Sequence 6, Applicant	4.30e-08

ALIGNMENTS

XX	RESULT	1
XX	PCT-US95-08493-13	STANDARD; PRT: 946 AA.
XX	xxxxxx	
DT		
DT		
DE	Sequence 13, Application PC/TUS9508493	
XX		
CC	Sequence 13, Application PC/TUS9508493	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Wood, Clive	
CC	APPLICANT: Caruso, Anthony	
CC	TITLE OF INVENTION: Novel mlk Receptor Tyrosine Kinases	
CC	NUMBER OF SEQUENCES: 21	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: LEGAL AFFAIRS	
CC	STREET: 87 Cambridgepark Drive	
CC	CITY: Cambridge	
CC	STATE: MA	
CC	COUNTRY: USA	
CC	ZIP: 02140	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: PCT/US95/08493	
CC	FILING DATE:	
CC	CLASSIFICATION:	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Brown, Scott A	
CC	REGISTRATION NUMBER: 32,724	
CC	REFERENCE/DOCKET NUMBER: G15234A	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (617) 498-8224	
CC	TELEFAX: (617) 876-5851	
CC	INFORMATION FOR SEQ ID NO: 13:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 946 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: protein	
CC	SEQUENCE 946 AA; 105892 MW; 4559778 CN;	
SO		

W096/b3

Query Match 33.9%; Score 207; DB 3; Length 946;


```
CC CC ATTORNEY/AGENT INFORMATION:  
CC CC NAME: OBLON, NORMAN F.  
CC CC REGISTRATION NUMBER: 24,618  
CC CC REFERENCE/DOCKET NUMBER: 18-966-0  
CC CC TELECOMMUNICATION INFORMATION:  
CC CC TELEPHONE: 703-413-3000  
CC CC TELEFAX: 703-413-2220  
CC CC INFORMATION FOR SRO ID NO: 1:  
CC CC SEQUENCE CHARACTERISTICS:  
CC CC LENGTH: 347 amino acids  
CC CC type: amino acid  
CC CC STRANDEDNESS: single  
CC CC TOPOLOGY: linear  
CC CC MOLECULE TYPE: protein  
CC CC SEQUENCE 347 AA; 38594 MW; 612249 CN;  
DB DB  
QY QY 20 SLTESGASCLPWNMSILGKVTYAQNPSAQLGLGKHNYCRNPDGAKPMWCHVLRN-RL 78  
QY 48 SVTFGAPCLRWAEVPFLER-SPSASQAQLRGO-RHNFCRSPDGARPMCFYDARGKV 105  
Db Db 79 TWGYCD 84  
QY 106 DWGYCD 111  
  
RESULT 5 Score 179; DB 2; Length 347;  
ID US-08-811-949-61 STANDARD: PRT; 354 AA.  
AC xxxxxx  
DT  
XX Sequence 61, Application US/08811949  
DE  
XX  
CC Sequence 61, Application US/08811949  
CC Patent No. 5840533  
CC GENERAL INFORMATION:  
CC APPLICANT: NIMA, MINEO  
CC APPLICANT: SATTO, YOSHIMASA  
CC APPLICANT: SASAKI, HITOSHI  
CC APPLICANT: HAYASHI, MASAO  
CC APPLICANT: NOTANI, JOUJI  
CC APPLICANT: KOBAYASHI, MASARAZU  
CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
CC NUMBER OF SEQUENCES: 67  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
CC ADDRESSEE: P.C.  
CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CC CITY: ARLINGTON  
CC STATE: VA  
CC COUNTRY: USA  
CC ZIP: 22202  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/811,949  
CC FILING DATE: 05-MAR-1997  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: OBLON, NORMAN F.  
CC REGISTRATION NUMBER: 24,618  
CC REFERENCE/DOCKET NUMBER: 18-966-0  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 703-413-3000  
CC TELEFAX: 703-413-2220  
CC
```

```






CC      INFORMATION FOR SEQ ID NO: 61:
CC      SEQUENCE CHARACTERISTICS:
CC          LENGTH: 354 amino acids
CC          TYPE: amino acid
CC          TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 354 AA; 39339 MW; 636420 CN;
Cc
Cc      Query Match          29.3%; Score 179; DB 2; Length 354;
Cc      Best Local Similarity 37.9%; Pred. No. 2,77e-08;
Cc      Matches 25; Conservative 15; Mismatches 23; Indels 3; Gaps 3;
Cc
Db      21 SITESGASCLPMNMSMILIGKYVTQNQNSAQQLGLGKKNYCNPCGDKAPMCHYLKNR-RL 79
Oy      48 SVTFEGACFLHMAEVPPLER-SEPASWAQLRGQ-RNHFCSRPGAGRPWCYGDARGKV 105
Db      80 TWGYCD 85
Oy      106 DMGYCD 111
Cc
Cc      RESULT 6
Cc      ID US-08-811-949-59 STANDARD; PRT: 355 AA.
Cc      XX xxxxxx
Cc      DT
Cc      XX
Cc      DE Sequence 59, Application US/08811949
Cc      CC Sequence 59, Application US/08811949
Cc      CC Patent No. 5840533
Cc      CC GENERAL INFORMATION:
Cc      CC APPLICANT: NIMA, MINEO
Cc      CC APPLICANT: SAITO, YOSHIMASA
Cc      CC APPLICANT: SASAKI, HITOSHI
Cc      CC APPLICANT: HAYASHI, MASAKO
Cc      CC APPLICANT: NOTANI, JOUJI
Cc      CC APPLICANT: KOBAYASHI, MASAAZU
Cc      CC TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
Cc      CC NUMBER OF SEQUENCES: 67
Cc      CC CORRESPONDENCE ADDRESS:
Cc      CC ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
Cc      CC ADDRESSEE: P.C.
Cc      CC STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
Cc      CC CITY: ARLINGTON
Cc      CC STATE: VA
Cc      CC COUNTRY: USA
Cc      CC ZIP: 22202
Cc      CC COMPUTER READABLE FORM
Cc      CC MEDIUM TYPE: Floppy disk
Cc      CC COMPUTER: IBM PC compatible
Cc      CC OPERATING SYSTEM: PC-DOS/MS-DOS
Cc      CC SOFTWARE: PatentIn Release #1.0, Version #1.30
Cc      CC CURRENT APPLICATION DATA:
Cc      CC APPLICATION NUMBER: US/08/811,949
Cc      CC FILING DATE: 05-MAR-1997
Cc      CC CLASSIFICATION: 435
Cc      CC ATTORNEY/AGENT INFORMATION:
Cc      CC NAME: OBLON, NORMAN F.
Cc      CC REGISTRATION NUMBER: 24,618
Cc      CC REFERENCE/DOCKET NUMBER: 18-966-0
Cc      CC TELECOMMUNICATION INFORMATION:
Cc      CC TELEPHONE: 703-413-3000
Cc      CC TELEFAX: 703-413-2220
Cc      CC INFORMATION FOR SEQ ID NO: 59:
Cc      CC SEQUENCE CHARACTERISTICS:
Cc      CC LENGTH: 355 amino acids
Cc      CC TYPE: amino acid
Cc      CC TOPOLOGY: linear
Cc      CC MOLECULE TYPE: protein
Cc      CC SEQUENCE 355 AA; 39350 MW; 636607 CN;
Cc
Cc      SEQUENCE 355 AA; 39350 MW; 636607 CN;
Cc

```


XX DE Sequence 2, Application US/08427640
 XX CC Sequence 2, Application US/08427640
 CC CC Patent No. 5658788
 CC CC GENERAL INFORMATION:
 CC CC APPLICANT: Berg et al.
 CC CC TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disc
 CC CC NUMBER OF SEQUENCES: 28
 CC CC CORRESPONDENCE ADDRESS:
 CC CC ADDRESSEE: Eli Lilly and Company
 CC CC STREET: Lilly Corporate Center
 CC CC CITY: Indianapolis
 CC CC STATE: IN.
 CC CC COUNTRY: U.S.A.
 CC CC ZIP: 46285
 CC CC COMPUTER READABLE FORM:
 CC CC MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
 CC CC COMPUTER: Macintosh
 CC CC OPERATING SYSTEM: Macintosh
 CC CC SOFTWARE: Microsoft Word
 CC CC CURRENT APPLICATION DATA:
 CC CC APPLICATION NUMBER: US/08/427,640
 CC CC FILING DATE:
 CC CC CLASSIFICATION: 435
 CC CC PRIOR APPLICATION DATA:
 CC CC APPLICATION NUMBER: 07/689,410
 CC CC FILING DATE: 22 APRIL 1991
 CC CC INFORMATION FOR SEQ ID NO: 2:
 CC CC SEQUENCE CHARACTERISTICS:
 CC CC LENGTH: 355 amino acids
 CC CC TYPE: amino acid
 CC CC TOPOLOGY: linear
 CC CC MOLECULE TYPE: Protein
 CC CC SEQUENCE 355 AA; 39589 MW; 639067 CN;
 SQ
 Query Match 29.3%; Score 179; DB 1; Length 355;
 Best Local Similarity 37.9%; Pred. No. 2.77e-08;
 Matches 25; Conservative 15; Mismatches 23; Indels 3; Gaps 3;
 DB 21 SLTESGASCLPWNMSMLIGKVTYTAQNSAOLGLGKHNCRNPDGDAKFWCHYLKNR-RL 79
 QY 48 SYTDFGAPCLRNAEVPFLLER-SPPASWQLRGO-RHNFCRSPDGAGRPWCYGDARGKV 105
 DB 80 TWGYCD 85
 QY 106 DWGYCD 111
 DB 106 DWGYCD 111
 ID 15
 XX US-08-427-640-6 STANDARD; PRT: 355 AA.
 XX AXXXXX
 DE Sequence 6, Application US/08427640
 XX CC Sequence 6, Application US/08427640
 CC CC Patent No. 5658788
 CC CC GENERAL INFORMATION:
 CC CC APPLICANT: Berg et al.
 CC CC TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disc
 CC CC NUMBER OF SEQUENCES: 28
 CC CC CORRESPONDENCE ADDRESS:
 CC CC ADDRESSEE: Eli Lilly and Company
 CC CC STREET: Lilly Corporate Center
 CC CC CITY: Indianapolis
 CC CC STATE: IN.
 CC CC COUNTRY: U.S.A.
 CC CC ZIP: 46285
 CC CC COMPUTER READABLE FORM:
 CC CC MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage

CC CC COMPUTER: Macintosh
 CC CC OPERATING SYSTEM: Macintosh
 CC CC SOFTWARE: Microsoft Word
 CC CC CURRENT APPLICATION DATA:
 CC CC APPLICATION NUMBER: US/08/427,640
 CC CC FILING DATE:
 CC CC CLASSIFICATION: 435
 CC CC PRIOR APPLICATION DATA:
 CC CC APPLICATION NUMBER: 07/689,410
 CC CC FILING DATE: 22 APRIL 1991
 CC CC INFORMATION FOR SEQ ID NO: 6:
 CC CC SEQUENCE CHARACTERISTICS:
 CC CC LENGTH: 355 amino acids
 CC CC TYPE: amino acid
 CC CC TOPOLOGY: linear
 CC CC MOLECULE TYPE: Protein
 CC CC SEQUENCE 355 AA; 39563 MW; 638774 CN;
 SQ
 Query Match 29.3%; Score 179; DB 1; Length 355;
 Best Local Similarity 37.9%; Pred. No. 2.77e-08;
 Matches 25; Conservative 15; Mismatches 23; Indels 3; Gaps 3;
 DB 21 SLTESGASCLPWNMSMLIGKVTYTAQNSAOLGLGKHNCRNPDGDAKFWCHYLKNR-RL 79
 QY 48 SYTDFGAPCLRNAEVPFLLER-SPPASWQLRGO-RHNFCRSPDGAGRPWCYGDARGKV 105
 DB 80 TWGYCD 85
 QY 106 DWGYCD 111
 Search completed: Mon Mar 13 10:23:02 2000
 Job time : 6 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Mon Mar 13 10:22:03 2000;      MasPar time 8.65 Seconds
Tabular output not generated.              398.104 Million cell updates/sec
```

```

Title: >US-09-147-947-6
Description: (40-112) from US09147947A pep (2 of 6)
Perfect Score: 611
Sequence: 1 CPAGPWNVSVIDFGAPLCRM.....RPFQFYDANGKVDWGICDC 73

```

Scoring table:

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

```
Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 36.459; Variance 65.448; scale 0.557

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length	DB	ID	Description	Prod. No.
1	507	83.0	761	2	JC5759	brain-specific serine	2.32e-88
2	207	33.9	946	1	AA7299	ror-related receptor	3.02e-23
3	185	30.3	615	1	KRH012	coagulation factor XI	7.43e-19
4	179	29.3	659	1	AA9941	t-plasminogen activator	1.12e-17
5	179	29.3	562	1	UKHUR	t-plasminogen activat	1.12e-17
6	175	28.6	716	1	AA0332	macrophage-stimulat	6.75e-17
7	174	28.5	812	1	PJBO	plasmin (EC 3.4.21.7)	1.06e-16
8	173	28.3	716	1	JC5061	macrophage-stimulat	1.65e-16
9	172	28.2	812	1	PJMS	plasmin (EC 3.4.21.7)	2.58e-16
10	171	28.0	603	2	S28941	coagulation factor XI	4.04e-16
11	171	28.0	790	1	PJPG	plasmin (EC 3.4.21.7)	4.04e-16
12	169	27.7	593	2	S45281	coagulation factor XI	9.83e-16
13	167	27.3	560	1	JC4795	plasma hyaluronan-bi	2.39e-15
14	165	27.0	558	2	JC5878	plasma hyaluronan-bi	5.78e-15
15	164	26.8	559	1	A35029	t-plasminogen activat	8.98e-15
16	163	26.7	728	1	UH0579	hepatocyte growth fac	1.40e-14
17	161	26.4	711	1	AA7136	macrophage-stimulat	3.36e-14
18	161	26.4	728	1	AA5644	hepatocyte growth fac	3.36e-14
19	161	26.4	728	1	AA0185	hepatocyte growth fac	3.36e-14
20	161	26.4	810	2	IA6260	plasmin (EC 3.4.21.7)	3.36e-14
21	159	26.0	710	1	II1283	hepatocyte growth fac	8.07e-14
22	157	25.7	4548	1	S00657	apoptotain(a) (EC 3.4	1.93e-13
23	156	25.5	169	2	AA0522	plasmin (EC 3.4.21.7)	2.99e-13

Use mit Abg. 2

Applicants

45	125	20.5	625	1	TBBO	thrombin (EC 3.4.21.5
44	126	20.1	617	2	UBBA	thrombin (EC 3.4.21.5
43	129	20.6	433	1	UBBA	u-plasminogen activat
42	131	21.4	618	2	UBBA	thrombin (EC 3.4.21.5
41	132	21.6	622	1	TBBU	thrombin (EC 3.4.21.5
40	132	21.6	291	2	138058	u-plasminogen activat
39	134	21.9	433	1	U00560	Plasmin (EC 3.4.21.7)
38	135	22.1	120	2	EA1545	Plasmin (EC 3.4.21.7)
37	137	22.4	455	2	EA1545	u-plasminogen activat
36	138	22.6	442	1	UKFP	u-plasminogen activat
35	139	22.6	431	1	JSU060	u-plasminogen activat
34	139	22.7	394	2	JSU060	u-plasminogen activat
33	140	22.9	433	1	UMMS	u-plasminogen activat
32	140	22.9	89	2	AE0140	u-plasminogen activat
31	142	23.2	432	1	518932	u-plasminogen activat
30	146	23.9	460	2	BE1545	Plasmin (EC 3.4.21.7)
29	146	23.9	123	2	C61545	Plasmin (EC 3.4.21.7)
28	147	24.1	810	1	PLHU	Plasmin (EC 3.4.21.7)
27	149	24.4	810	2	B30848	heparocyte growth fac
26	149	24.4	411	2	B30848	heparocyte growth fac
25	150	24.5	655	1	A65688	heparocyte growth fac
24	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
23	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
22	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
21	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
20	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
19	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
18	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
17	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
16	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
15	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
14	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
13	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
12	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
11	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
10	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
9	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
8	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
7	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
6	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
5	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
4	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
3	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
2	154	25.2	1420	2	A23869	apolipoprotein(a) (EC
1	154	25.2	1420	2	A23869	apolipoprotein(a) (EC

ALIGNMENTS

```

ENTRY          #type complete
TITLE          brain-specific serine proteinase (EC 3.4.21.-) - mouse
ORGANISM       Mus musculus #common_name mouse
DATE           24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change
               17-Mar-1999
ACCESSIONS     JC5759
REFERENCE      JC5759
#authors       Yamamura, Y.; Yamashiro, K.; Tsuruoka, N.; Nakazato, H.;
               Tsujimura, A.; Yanaguchi, N.
               #journal      Biochem. Biophys. Res. Commun. 1997, 239:386-392
               #title       Molecular cloning of a novel brain-specific serine proteinase
               with a kringle-like structure and three scavenger receptor
               cysteine-rich motifs.
#cross-references MIMD:98008948
#accession     JC5759
#molecule_type mRNA
#residues      1-761 #label YAM
#cross-references DDBJ:D89871
#experimental_source brain
CLASSIFICATION #superfamily trypsin homology; scavenger receptor
               cysteine-rich domain homology
               glycoprotein; hydrolase; serine proteinase
KEYWORDS        #domain kringle-like #status predicted #label KRI\
FEATURE          #domain scavenger receptor cysteine-rich domain homology
               #label SRC\
               163-266
               85-157
               163-266
               166-266, 273-372,
               366-486
               #domain scavenger receptor cysteine-rich #status
               predicted #label SRC\
               513-516
               517-755
               93,521,569
               #domain furin binding #status predicted #label FRB\
               #domain trypsin homology #label TRY\
               #binding_site carbohydrate (asn) (covalent) #status
               predicted\
               562,612,711
               #active_site His, Asp, Ser #status predicted
SUMMARY         #length 761 #molecular_weight 84136 #checksum 5449
               83.0%: Score 507, DB 2: Length 761;
Query Match     Pred Local Similarity 78.1%:   Pred. No.2,32e-88;
Matches         57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;
Db              85 CGAGESWGNATNLGVPCLHWDEVPLPFLERSPPASNAELRGPHNFCRSPDSGRPWCFYR 144
               I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::
Oy              40 CPAGEPMVSTYDFGAPCLRWAEVPLPFLERSPPASNAQLRGQHNCRCRSPDAGRWCFYTG 99
               I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I::
Db              145 NAAGKVDWGYCDD 157
               I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I:::I::

```

QY 100 DARGVDMGYDCD 112

RESULT 2

ENTRY A47299 #type complete

TITLE ror-related receptor Rrk - Pacific electric ray

CONTAINS protein-tyrosine kinase (EC 2.7.1.112)

ORGANISM #formal_name Torpedo californica #common_name Pacific electric ray

DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS A47299

REFERENCE A47299

#authors Jennings, C.G., Dyer, S.M., Burden, S.J., Proc. Natl. Acad. Sci. U.S.A. (1993) 90:2895-2899

#journal Muscle-specific rkr-related receptor with a kinase domain defines a distinct class of receptor tyrosine kinases.

#title

#cross-references MUID:9219391

Accession A47299

#status Preliminary

#molecule_type mRNA

#residues 1-946 #label JEN

#cross-references GB:111311; NID:9290857; PIDN:AAA49285.1; PID:9290858

#experimental_source electric organ

#note sequence extracted from NCBI backbone (NCBIN:128724, NCBI:128726)

CLASSIFICATION #superfamily Torpedo ror-related receptor; kringle homology; protein kinase homology

KEYWORDS ATP; glycoprotein; kringle; phosphotransferase; transmembrane protein; tyrosine-specific protein kinase

FEATURE

464-542 #domain kringle homology #label KR3\

572-588 #domain transmembrane #status predicted #label TMN\

650-940 #domain protein kinase homology #label KIN\

658-666 #region protein kinase ATP-binding motif\

225,340,477,544 #binding_site carbohydrate (asn) (covalent) #status predicted

SUMMARY #length 946 #molecular_weight 105892 #checksum 8717

Query Match 33.9%; Score 207; DB 1; Length 946;

Best Local Similarity 39.4%; Pred. No. 3,026-23;

Matches 26; Conservative 13; Mismatches 23; Indels 4; Gaps 4;

DB 476 VNTASISICORSEQAPHFHRRLPELPPEL-ANSDNFCRNGGSEPPWC-YTMDRP-I 532

QY 47 VSYTDFCAPLRMAEVPFLERSPASWQDRGQRHNCRSPDGAG-RPWCFTGDARGKV 105

533 RMEFCN 538

106 DWGYCD 111

ENTRY 3

TITLE KFMU12 #type complete

ALTERNATE_NAMES coagulation factor XIIa (EC 3.4.21.38) precursor - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 27-Nov-1985 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

ACCESSIONS A29411; A26814; A00930; A25191; A22248; A21037

REFERENCE A29411

#authors Cool, D.E.; MacGillivray, R.T.A., J. Biol. Chem. (1987) 262:13662-13673

#journal Characterization of the human blood coagulation factor XII gene, Intron/exon gene organization and analysis of the 5'-flanking region.

#cross-references MUID:88007593

Accession A29411

#molecule_type DNA

#residues 1-615 #label COO

#cross-references GB:M17466; GB:J02807; NID:9180355; PIDN:AA59490.1; PID:9180357

REFERENCE A26814

#authors Tripodi, M.; Citarella, F.; Guida, S.; Galeffi, P.; Fantoni, A.; Correse, R.

#journal Nucleic Acids Res. (1986) 14:3146

#title cDNA sequence coding for human coagulation factor XII (Hageman).

#cross-references MUID:86176794

Accession A26814

#molecule_type mRNA

#residues 4-615 #label TRI

#cross-references GB:M13135; NID:9182291; PIDN:AAA70225.1; PID:9182292

REFERENCE A00930

#authors Cool, D.E.; Edgell, C.J.S.; Louie, G.V.; Zoller, M.J.; Brayer, G.D.; MacGillivray, R.T.A., J. Biol. Chem. (1985) 260:13666-13676

#journal Characterization of human blood coagulation factor XII cDNA. Prediction of the primary structure of factor XII and the tertiary structure of beta-factor XIIa.

#cross-references MUID:86033830

Accession A00930

#molecule_type mRNA

#residues 14-332, 'S', 334-615 #label CO2

#cross-references GB:M11723; NID:9180358; PIDN:AAA51986.1; PID:9180359

REFERENCE A25191

#authors Que, B.G.; Davie, E.W., Biochemistry (1986) 25:1525-1528

#journal Characterization of a cDNA coding for human factor XII (Hageman factor).

#cross-references MUID:86216049

Accession A25191

#molecule_type mRNA

#residues 146-378, 'G', 380-615 #label OUE

#cross-references GB:M3147; NID:9180360; PIDN:AAA70224.1; PID:9180361

REFERENCE A22248

#authors McMullen, B.A.; Fujikawa, K., J. Biol. Chem. (1985) 260:5328-5341

#journal Amino acid sequence of the heavy chain of human alpha-factor XIIa (activated Hageman factor).

#cross-references MUID:85182674

Accession A22248

#molecule_type protein

#residues 20-379 #label MCM

REFERENCE A21037

#authors Fujikawa, K.; McMullen, B.A., J. Biol. Chem. (1985) 258:10924-10933

#journal Amino acid sequence of human beta-factor XIIa.

#cross-references MUID:83291041

Accession A21037

#molecule_type protein

#residues 354-362; 373-615 #label FUJ

REFERENCE A44606

#authors Harris, R.J.; Ling, V.T.; Spellman, M.W., J. Biol. Chem. (1992) 267:5102-5107

#journal O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C.

#cross-references MUID:92184750

#contants annotation; carbohydrate binding site

GENETICS

#gene GDB:F12

#cross-references GDB:119892; OMIM:234000

#map_position 5q34-5qter

#introns 19/3; 39/1; 72/2; 96/1; 133/1; 177/1; 212/1; 267/2; 340/1; 417/2; 463/1; 511/1; 560/3

COMPLEX

#description factor XIIa catalyzes the proteolytic activation of plasminogen, plasma prekallikrein, and coagulation factors VII and IX; factor XII can catalyze the proteolytic activation of prekallikrein

FUNCTION

#description blood coagulation; fibrinolysis

CLASSIFICATION

#pathway #superfamily coagulation factor XII; EGF homology; fibronectin type I repeat homology; fibronectin type II repeat homology; kringle homology; trypsin homology


```

#molecule-type DNA
#residues 1-562 ##label NYT
#cross-references GB:L00141
#note the codon given for residue 93 (ACC) is inconsistent
      with the authors' translation
REFERENCE
#authors A23529
#journal Friezner Degen, S.J.; Rajput, B.; Reich, E.
#journal J. Biol. Chem. (1986) 261:6972-6985
#title The human tissue plasminogen activator gene.
#cross-references MIM:86196143
#accession A23529
##molecule-type DNA
##residues 1-562 ##label DEG
#cross-references GB:R03021; NID:q339817; PIDN:AAA98809.1; PID:q339818
#accession J10562
#authors Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.
#journal Agric. Biol. Chem. (1991) 55:1225-1232
#title Purification and characterization of tissue plasminogen
      activator secreted by human embryonic lung diploid
      fibroblasts, IMR-90 cells.
#cross-references MIM:91291340
#accession J10562
##molecule-type mRNA
##residues 31-562 ##label ITA
#cross-references DDBJ:D01096; NID:q220128; PIDN:BA00881.1;
      PID:d1001345; PID:g441174
#experimental_source embryonic lung fibroblast IMR-90 cells
#note mature protein, was confirmed by protein sequencing
REFERENCE
#authors A93293
#journal Penlica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar,
      G.A.; Ward, C.A.; Bennett, W.F.; Yelverton, E.; Seeburg,
      P.H.; Heyneker, H.L.; Goeddel, D.V.; Collen, D.
#journal Nature (1983) 301:214-221
#title Cloning and expression of human tissue-type plasminogen
      activator cDNA in Escherichia coli.
#cross-references MIM:83115262
#accession A93293
##molecule-type mRNA
##residues 1-562 ##label PEN
#cross-references GB:L00141
#experimental_source melanoma cells
REFERENCE
#authors S02125
#journal Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.
#journal Nucleic Acids Res. (1988) 16:5695
#title Nucleotide sequence of the tissue-type plasminogen activator
      cDNA from human fetal lung cells.
#cross-references MIM:88262579
#accession S02125
##status translation not shown
##molecule-type mRNA
##residues 1-562 ##label SAS
#cross-references EMBL:X07393; NID:q37243; PIDN:CA030302.1; PID:g37244
#experimental_source fetal lung cells
REFERENCE
#authors A91343
#journal Kagitani, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.;
      Bando, H.; Okada, K.; Matsuo, O.
#journal FEBS Lett. (1985) 189:145-149
#title Expression in Escherichia coli of finger-domain lacking
      tissue-type plasminogen activator with high fibrin
      affinity.
#cross-references MIM:85285620
#accession A91343
##molecule-type mRNA
##residues 1-38, 'G', 86-433 'E', 435-562 ##label KAG
#experimental_source Detroit 562 cells; ATCC 138
REFERENCE
#authors A93951
#journal Edlund, T.; Ny, T.; Ranby, M.; Hedén, L.O.; Palm, G.;
      Holmgren, E.; Josephson, S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1983) 80:349-352
#title Isolation of cDNA sequences coding for a part of human tissue
      plasminogen activator.
#cross-references MIM:8316956
#accession A93951
##molecule-type mRNA
##residues 251-358 ##label EDL
#experimental_source melanoma cells
REFERENCE
#authors A90488
#journal Pohl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jorvall, H.
#journal Biochemistry (1984) 23:3701-3707
#title Tissue plasminogen activator: peptide analyses confirm an
      indirectly derived amino acid sequence, identify the active
      site serine residue, establish glycosylation sites, and
      localize variant differences.
#cross-references MIM:85000468
#accession A91322
#contents annotation; melanoma cells, partial sequence of residues
      36-562, active and binding sites, heterogeneity
#cross-references MIM:84158956
#accession A91322
##molecule-type protein
##residues 33-45;311-320 ##label POH
#experimental_source uterus
#note in the uterus, cleavage of the activation peptide may
      also occur after 38-Gln
REFERENCE
#authors A37567
#journal van Zonneveld, A.J.; Veerman, H.; Pannekoek, H.
#journal J. Biol. Chem. (1986) 261:14214-14218
#contents annotation; fibrin binding site
REFERENCE
#authors A37568
#journal Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munk,
      G.A.W.; Pouwels, P.H.; Enger-Valk, B.E.
#journal EMBO J. (1986) 5:3525-3530
#title Involvement of finger domain and kringle 2 domain of
      tissue-type plasminogen activator in fibrin binding and
      stimulation of activity by fibrin.
#cross-references MIM:87161761
#accession A60902
#contents annotation; fibrin binding site
REFERENCE
#authors A60902
#journal Dodd, I.; Nunn, B.; Robinson, J.H.
#journal Thromb. Haemost. (1988) 59:523-528
#title Isolation, identification and pharmacokinetic properties of
      human tissue-type plasminogen activator species: possible
      localisation of a clearance recognition site.
#cross-references MIM:89044681
#accession A54645
#contents annotation; novel forms of expressed recombinant t-PA
REFERENCE
#authors A54645
#journal Harris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.;
      Emtage, J.S.; Odenakker, G.; Volckert, G.; Romants, W.;
      Billiau, A.; De Somer, P.
#journal Mol. Biol. Med. (1986) 3:279-292
#title Cloning of cDNA coding for human tissue-type plasminogen
      activator and its expression in Escherichia coli.
#cross-references MIM:86284200
#accession A54645
##molecule-type mRNA
##residues 1-362 ##label HAR
#cross-references GB:M15518; NID:q190031; PIDN:AAA6011.1; PID:q190032
#note parts of this sequence were confirmed by peptide
      sequencing
REFERENCE
#authors I60110
#journal Reddy, V.B.; Garramone, A.J.; Sasaki, H.; Wei, C.
#journal DNA (1987) 6:461-472
#title Expression of human uterine tissue-type plasminogen activator
      in mouse cells using BPV vectors.
#cross-references MIM:88054470
#accession I60110
##status translated from GB/EMBL/DDBJ
##molecule-type mRNA

```



```

30-308      #product t-plasminogen activator chain A #status
38-75      #predicted #label ACH\
83-116      #domain fibronectin type I repeat homology #label 1F1\
124-205      #domain EGF homology #label EGF\
213-294      #domain kringle homology #label KR1\
303-333      #domain kringle homology #label KR2\
309-553      #product t-plasminogen activator chain B #status
38-68,66-75,83-94,  #predicted #label BCH\
88-105,107-116,  #domain trypsin homology #label TRY\
124-205,145-187,
176-200,213-284,
234-276,265-289,
297-428,340-356,
348-417,442-516,
474-490,506-534,
149,481      #disulfide bonds #status predicted\
309-309      #binding site carbohydrate (asn) (covalent) #status
              #predicted\
              #cleavage_site Arg-Ile (plasmin, trypsin) #status
              #predicted\
              #active_site His, Asp, Ser #status predicted
SUMMARY      #length 559 #molecular_weight 62903 #checksum 8573

Query Match      26.8%: Score 164; DB 1; Length 559;
Best Local Similarity 41.0%: Pred.No.8,98e-15;
Matches 25: Conservative 12; Mismatches 19; Indels 5; Gaps 4;

DB      232 ASCLPWNMSI-LIGITYAWRANSQALGLGRHNYCRNPDGDAKPCWCHMKDR-KLTWEYC 289
QY      54 APLRLMAEVPPEFLERSPPASWA--QLRGO-RHNFCRSPDGAGRPCFYGDARGKVDWGXC 110

DB      290 D 290
QY      111 D 111

```

Search completed: Mon Mar 13 10:22:13 2000
 Job time : 10 secs.

 NWSENF
 ***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:21:05 2000; Maspar time 5.46 Seconds
 Tubular output not generated. 399,446 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (40-112) from US09147947A.pep (2 of 6)
 Perfect Score: 611
 Sequence: 1 CPAGEPWSVTDFGAPCLRW.....RPMCFYGDARGKVDWGYCDC 73

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: Swiss-prot33
 I:swissprot

Statistics: Mean 37.173; Variance 60.386; scale 0.616

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
611	100.0	875	1	NETR_HUMAN	1.14e-123
507	83.0	761	1	NETR_MOUSE	1.24e-97
185	30.3	615	1	FA12_HUMAN	4.54e-21
119	29.3	559	1	UROT_MOUSE	8.93e-20
179	29.3	562	1	UROT_HUMAN	6.42e-19
175	28.6	716	1	HGFL_MOUSE	1.05e-18
174	28.2	812	1	PLMN_BOVIN	2.80e-18
172	28.2	812	1	PLMN_MOUSE	4.57e-18
171	28.0	603	1	FA12_CAVPO	4.57e-18
111	28.0	790	1	PLMN_PTG	1.24e-17
112	27.7	593	1	FA12_BOVIN	1.38e-16
169	26.8	559	1	UROT_RAT	2.23e-16
163	26.7	728	1	HGF_HUMAN	5.86e-16
161	26.4	711	1	HGFL_HUMAN	5.86e-16
161	26.4	728	1	HGF_MOUSE	5.86e-16
161	26.4	728	1	HGF_RAT	5.86e-16
161	26.4	810	1	PLMN_ERIEU	3.99e-15
157	25.7	4548	1	APOL_HUMAN	6.43e-15
156	25.5	169	1	PLMN_RAT	1.67e-14
154	25.2	1420	1	APOL_MOUSE	1.11e-13
150	24.5	655	1	HGFL_HUMAN	1.78e-13
149	24.4	810	1	PLMN_MOUSE	4.54e-13
147	24.1	810	1	PLMN_HUMAN	4.54e-13

Usefuls 243

RESULT	1	STANDARD	PRT	875 AA
ID	NETR_HUMAN			
AC	P56730			
DT	15-DEC-1999 (Rel. 39, Created)			
DT	15-DEC-1999 (Rel. 39, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	NEUTROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin).			
GN	PRSS12			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN:			
RA	MEDLINE: 98201705.			
RT	PROBA K., GSCHEWEND T.P., SONDEREGGER P.;			
RL	"Cloning and sequencing of the cDNA encoding human neutrotrypsin.";			
CC	Biochim. Biophys. Acta 1396:143-147 (1998).			
CC	- FUNCTION: PLAYS A ROLE IN NEUTROTRYPsin ACTIVITY AND THE PROTEOLYTIC			
CC	ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH			
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.			
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/			
CC	or send an email to license@sdb-sdb.ch).			
CC	EMBL: AJ001531; CAA04816.1; -			
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.			
DR	PROSITE: PS00135; TRYPSIN_SER. 1.			
DR	PROSITE: PS00420; SERPACT_RECEPTOR. 3.			
DR	HydroLase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	875	NEUTROTRYPsin.
FT	DOMAIN	23	92	PROLINE-RICH.
FT	DOMAIN	93	165	KRINGLE.
FT	DOMAIN	170	271	SRCR 1.
FT	DOMAIN	170	280	SRCR 2.
FT	DOMAIN	387	487	SRCR 3.

ALIGNMENTS

ID	NETR_HUMAN	STANDARD	PRT	875 AA
AC	P56730			
DT	15-DEC-1999 (Rel. 39, Created)			
DT	15-DEC-1999 (Rel. 39, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	NEUTROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin).			
GN	PRSS12			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN:			
RA	MEDLINE: 98201705.			
RT	PROBA K., GSCHEWEND T.P., SONDEREGGER P.;			
RL	"Cloning and sequencing of the cDNA encoding human neutrotrypsin.";			
CC	Biochim. Biophys. Acta 1396:143-147 (1998).			
CC	- FUNCTION: PLAYS A ROLE IN NEUTROTRYPsin ACTIVITY AND THE PROTEOLYTIC			
CC	ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH			
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.			
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/			
CC	or send an email to license@sdb-sdb.ch).			
CC	EMBL: AJ001531; CAA04816.1; -			
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.			
DR	PROSITE: PS00135; TRYPSIN_SER. 1.			
DR	PROSITE: PS00420; SERPACT_RECEPTOR. 3.			
DR	HydroLase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	875	NEUTROTRYPsin.
FT	DOMAIN	23	92	PROLINE-RICH.
FT	DOMAIN	93	165	KRINGLE.
FT	DOMAIN	170	271	SRCR 1.
FT	DOMAIN	170	280	SRCR 2.
FT	DOMAIN	387	487	SRCR 3.

```

FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE, PROTEASE.
FT ACT_SITE 619 630 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67D5272B CRC32:

Query Match
Best Local Similarity 100.0%; Score 611; DB 1; Length 875;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 93 CPAGEPWSVTDFGAPCLRWAEVPEFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 152
40 CPAGEPWSVTDFGAPCLRWAEVPEFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 99
153 DARGKVMGXCDC 165
100 DARGKVMGXCDC 112

RESULT 2
ID NEUR_MOUSE STANDARD: PRT: 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUROTRYPsin, a novel multidomain serine protease expressed in the
DE PROTEASE 3) (BSSP-3).
GN PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GOSCHWEND T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neurotrypsin, a novel multidomain serine protease expressed in the
RT nervous system."
RT Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT single-like structure and three scavenger receptor cysteine-rich
RT motifs."
RT Biochem. Biophys. Res. Commun. 239:386-392(1997).
RN [1]
RP BLOCHM. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
CC LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
CC AND AMYGALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: Y13192; CAAT3646.1;
DR EMBL: D89871; BAA23986.1;
DR MGd; MG1:1100881; PRSS12.
DR PFAM: PF00530; SRCR; 3.
DR PFAM: PF00089; trypsin; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
KW Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761
FT DOMAIN 85 157
FT DOMAIN 166 267
FT DOMAIN 273 373
FT DOMAIN 386 487
FT DOMAIN 505 761
FT DOMAIN 505 516
FT ACT_SITE 516 517 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 562 562 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 CHARGE RELAY SYSTEM.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 84118 MW: 3F3CAF35 CRC32:

Query Match
Best Local Similarity 78.1%; Score 507; DB 1; Length 761;
Matches 57; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

Db 85 CGAESMGNAINLGVCLHMDVEVPEFLERSPPASMAELRGQRHNCRSPDGRPMCFY 144
40 CPAGEPWSVTDFGAPCLRWAEVPEFLERSPPASMAQLRGQRHNCRSPDGRPMCFY 99
145 NAQKVMGXCDC 157
100 DARGKVMGXCDC 112

RESULT 3
ID FA12_HUMAN STANDARD: PRT: 615 AA.
AC P00748;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COAGULATION FACTOR XII PRECURSOR (EC 3.4.21.38) (HAGEMAN FACTOR)
DE (HAF).
GN F12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88007593.
RA COOL D.E., MCGILLIVRAY R.T.A.;
RT "Characterization of the human blood coagulation factor XII gene.
RT Intron/exon gene organization and analysis of the 5'-flanking
RT region."
RT J. Biol. Chem. 262:13662-13673(1987).
RN [2]
RP SEQUENCE OF 4-615 FROM N.A.
RX MEDLINE: 86176794.
RA TRIPODI M., CITARELLA F., GUIDA S., GALEFFI P., FANTONI A.,
RA CORPSE R.;
RT "cDNA sequence coding for human coagulation factor XII (Hageman).";
RT Nucleic Acids Res. 14:3146-3146(1986).
RN [3]
RP SEQUENCE OF 14-615 FROM N.A.
RX MEDLINE: 8603830.
RA COOL D.E., ENGELL C.J.S., LOUIE G.V., ZOLLER M.J., BAYER G.D.,
RA MCGILLIVRAY R.T.A.;
RT "Characterization of human blood coagulation factor XII cDNA.

```

RT Prediction of the primary structure of factor XII and the tertiary
 RT structure of beta-factor XIIa.";
 RL J. Biol. Chem. 260:1366-1367(1985).
 RN [4]
 RP SEQUENCE OF 146-615 FROM N.A.
 RX MEDLINE: 86216049.
 RA OUE B.G., DAVIS E.W.;
 RT "Characterization of a cDNA coding for human factor XII (Hageman
 RI factor).";
 RL Biochemistry 25:1525-1528(1986).
 RN [5]
 RP SEQUENCE OF 20-379.
 RX MEDLINE: 85182674.
 RA MCMULLEN B.A., FUJIKAWA K.;
 RT "Amino acid sequence of the heavy chain of human alpha-factor XIIa
 (activated Hageman factor).";
 RL J. Biol. Chem. 260:5328-5341(1985).
 RN [6]
 RP SEQUENCE OF 354-362 AND 373-615.
 RX MEDLINE: 83291041.
 RA FUJIKAWA K., MCMULLEN B.A.;
 RT "Amino acid sequence of human beta-factor XIIa.";
 RL J. Biol. Chem. 258:10924-10933(1983).
 RN [7]
 RP VARIANT WASHINGTON DC.
 RX MEDLINE: 90045788.
 RA MIYATA T., KAWABATA S.-I., IWANAGA S., TAKAHASHI I., ALVING B.,
 RA SAITO H.;
 RT "Coagulation factor XII (Hageman factor) Washington D.C.: inactive
 RT factor XIIa results from Cys-571-->Ser substitution.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8319-8322(1989).
 RN [8]
 RP VARIANT LOCARNO.
 RX MEDLINE: 94325559.
 RA HOVINGA J.K., SCHALLER J., STRICKER H., WUILLEMIN W.A., FURLAN M.,
 RA LAMBLE B.;
 RT "Coagulation factor XII Locarno: the functional defect is caused by
 RT the amino acid substitution Arg-353-->Pro leading to loss of a
 RL kallikrein cleavage site.";
 RL Blood 84:1173-1181(1994).
 RN [9]
 RP CARBOHYDRATE-BINDING SITE THR-109.
 RX MEDLINE: 92184750.
 RA HARRIS R.J., LING V.T., SPELLMAN M.W.;
 RT "O-linked fucose is present in the first epidermal growth factor
 RT domain of factor XII but not protein C.";
 RL J. Biol. Chem. 267:5102-5107(1992).
 RN [10]
 RP FUNCTION: FACTOR XII IS A SERUM GLYCOPROTEIN THAT PARTICIPATES IN
 RT THE INITIATION OF BLOOD COAGULATION, FIBRINOLYSIS, AND THE
 RT GENERATION OF BRADYKININ AND ANGIOTENSIN.
 CC -1- CATALYTIC ACTIVITY: CLEAVES SELECTIVELY ARG-|-ILE BONDS AND
 CC -1- ACTIVATES COAGULATION FACTORS VII AND XI.
 CC -1- PTM: O- AND N-GLYCOSYLATED.
 CC -1- DISEASE: DEFECTS IN F12 DO NOT CAUSE ANY CLINICAL SYMPTOMS. THE
 CC -1- SOLE EFFECT IS THAT WHOLE-BLOOD CLOTTING TIME IS PROLONGED.
 CC -1- MISCELLANEOUS: FACTOR XII, PREKALLIKREIN, AND HMW KININOGEN FORM A
 CC -1- COMPLEX BOUND TO AN ANIONIC SURFACE. PREKALLIKREIN IS CLEAVED BY
 CC -1- FACTOR XII TO FORM KALLIKREIN, WHICH THEN CLEAVES FACTOR XII FIRST
 CC -1- TO ALPHA-FACTOR XIIA AND THEN TO BETA-FACTOR XIIA. ALPHA-FACTOR
 CC -1- XIIA ACTIVATES FACTOR XI TO FACTOR XIA.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE REGION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC -1- TRIPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: M31315; AAA70225.1; -;
 DR EMBL: M11723; AAA51986.1; -;
 DR EMBL: M17466; AAB59490.1; -;
 DR EMBL: M17464; AAB59490.1; JOINED.
 DR EMBL: M17465; AAB59490.1; JOINED.
 DR EMBL: M13147; AAA70224.1; -;
 DR PIR: A29411; KFH012.
 DR HSSP: P00763; IDPO.
 DR MIM: 234000; -;
 DR PROSITE: PS00021; KRINGLE_1; 1.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS00023; FIBRONECTIN_2; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01253; FIBRONECTIN_1; 1.
 DR PROSITE: PS0070; KRINGLE_2; 1.
 DR PFAM: PF00008; EGF; 2.
 DR PFAM: PF00039; fn1; 1.
 DR PFAM: PF00040; fn2; 1.
 DR PFAM: PF00051; kringle; 1.
 DR PFAM: PF00089; trypsin; 1.
 DR Glycoprotein; Blood coagulation; Plasma; kringle; Serine protease;
 DR Hydrolase; Fibrinolysis; Signal; EGF-like domain; Repeat; zymogen;
 DR Disease mutation.
 FT SIGNAL 1 19
 FT CHAIN 20 372 ALPHA-FACTOR XIIA HEAVY CHAIN.
 FT CHAIN 373 615 ALPHA-FACTOR XIIA LIGHT CHAIN.
 FT CHAIN 354 362 BETA-FACTOR XIIA PART 1.
 FT CHAIN 373 615 BETA-FACTOR XIIA PART 2.
 FT DOMAIN 47 88 FIBRONECTIN TYPE-II.
 FT DOMAIN 94 131 EGF-LIKE 1.
 FT DOMAIN 133 173 FIBRONECTIN TYPE-I.
 FT DOMAIN 174 210 EGF-LIKE 2.
 FT DOMAIN 217 295 KRINGLE.
 FT DOMAIN 296 349 PRO-RICH.
 FT DOMAIN 373 615 CATALYTIC.
 FT CARBOHYD 109 109 FUCOSE.
 FT CARBOHYD 249 249
 FT CARBOHYD 299 299 POTENTIAL.
 FT CARBOHYD 305 305 POTENTIAL.
 FT CARBOHYD 308 308 POTENTIAL.
 FT CARBOHYD 328 328 POTENTIAL.
 FT CARBOHYD 329 329 POTENTIAL.
 FT CARBOHYD 337 337 POTENTIAL.
 FT ACT_SITE 412 412 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 461 461 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 563 563 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 98 110 BY SIMILARITY.
 FT DISULFID 104 119 BY SIMILARITY.
 FT DISULFID 121 130 BY SIMILARITY.
 FT DISULFID 135 163 BY SIMILARITY.
 FT DISULFID 161 170 BY SIMILARITY.
 FT DISULFID 178 189 BY SIMILARITY.
 FT DISULFID 183 198 BY SIMILARITY.
 FT DISULFID 200 209 BY SIMILARITY.
 FT DISULFID 217 295 BY SIMILARITY.
 FT DISULFID 238 277 BY SIMILARITY.
 FT DISULFID 266 280 BY SIMILARITY.
 FT DISULFID 359 486 BY SIMILARITY.
 FT DISULFID 397 413 BY SIMILARITY.
 FT DISULFID 405 475 BY SIMILARITY.
 FT DISULFID 436 439 BY SIMILARITY.
 FT DISULFID 500 569 BY SIMILARITY.
 FT DISULFID 532 548 BY SIMILARITY.
 FT DISULFID 559 590 BY SIMILARITY.
 FT VARIANT 372 372 R -> P (IN LOCARNO; INACTIVE).
 FT VARIANT 372 372 C -> S (IN WASHINGTON DC; INACTIVE).
 FT VARIANT 590 590 /FTID=VAR_006623.
 FT VARIANT 590 590 /FTID=VAR_006624.
 FT CONFLICT 333 333 P -> S (IN REF. 3).

RX MEDLINE: 88262579.
RA SASAKI H., SAITO Y., HAYASHI M., OTSUKA K., NIMA M.;
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA
from human fetal lung cells.";
RL Nucleic Acids Res. 16:5695-5695(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88054470.
RA REDDY V.B., GARRAMONE A.J., SASAK H., WEI C.M., WATKINS P., GALLI J.,
RA HSUNG N.;
RT "Expression of human uterine tissue-type plasminogen activator in
mouse cells using BPV vectors.";
RL DNA 6:461-472(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86196143.
RA DEGEN S.J.F., RAJPUT B., REICH E.;
RT "The human tissue plasminogen activator gene.";
RL J. Biol. Chem. 261:6972-6985(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84298137.
RA NY T., ELGH F., LUND B.;
RT "The structure of the human tissue-type plasminogen activator gene:
correlation of intron and exon structures to functional and
structural domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86284200.
RA HARRIS T.J., PATEL T., MARSTON F.A., LITTLE S., EMRAGE J.S.,
RA OPDENAKKER G., VOLCKAERT G., ROMBAUTS W., BILLIAU A., SOMER P.;
RT "Cloning of cDNA coding for human tissue-type plasminogen activator
and its expression in Escherichia coli.";
RL Mol. Biol. Med. 3:279-292(1986).
RN [7]
RP SEQUENCE OF 251-358 FROM N.A.
RX MEDLINE: 83169656.
RA EDLUND T., NY T., RANBY M., HEDEN L.-O., PALM G., HOLMGREN E.,
RA JOSEPHSON S.;
RT "Isolation of cDNA sequences coding for a part of human tissue
plasminogen activator.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352(1983).
RN [8]
RP SEQUENCE OF 31-562 FROM N.A.
RX MEDLINE: 91291340.
RA ITAKAKI Y., YASUDA H., MORINAGA T., MITSUDA S., HIGASHIO K.;
RT "Purification and characterization of tissue plasminogen activator
secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";
RL Agric. Biol. Chem. 55:1225-1232(1991).
RN [9]
RP SEQUENCE OF 36-562.
RX MEDLINE: 85000468.
RA POHL G., KALLSTRÖM M., BERGSDORF N., WALLEN P., JOERNVALL H.;
RT "Tissue plasminogen activator: peptide analyses confirm an indirectly
derived amino acid sequence, identify the active site serine residue,
establish glycosylation sites, and localize variant differences.";
RL Biochemistry 23:3701-3707(1984).
RN [10]
RP SEQUENCE FROM N.A. (SMALL ISOFORM).
RX MEDLINE: 90192129.
RA SIEBERT P.D., FONG K.;
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from
human endothelial cells.";
RL Nucleic Acids Res. 18:1086-1086(1990).
RN [11]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE: 90092112.
RA PEIERFER G., SCHMIDT M., STRUBE K.-H., GEYER R.;
RT "Carbohydrate structure of recombinant human uterine tissue
plasminogen activator expressed in mouse epithelial cells.";
RL Eur. J. Biochem. 186:273-286(1989).

RN [12]
RP DISULFIDE BONDS IN KRINGLE 2.
RX MEDLINE: 91244765.
RA VLACHOS C.J., WILHELM O.G., HASSELT T., JASKUNAS S.R., BANG N.U.;
RT "Disulfide pairing of the recombinant kringle-2 domain of tissue
plasminogen activator produced in Escherichia coli.";
RL J. Biol. Chem. 266:10070-10072(1991).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.
RX MEDLINE: 96200985.
RA LABA D., BAUER M., HUBER R., FISCHER S., RUDOLPH R., KOHNERT U.,
RA BODE M.;
RT "The 2.3 Å crystal structure of the catalytic domain of recombinant
two-chain human tissue-type plasminogen activator.";
RL J. Mol. Biol. 258:117-135(1996).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.
RX MEDLINE: 92118803.
RA DE VOS A., ULTSCH M.H., KELLEY R.F., PADMANABHAN K., TULINSKI A.,
RA WESTBROOK M.L., KOSSIAKOF A.A.;
RT "Crystal structure of the kringle 2 domain of tissue plasminogen
activator at 2.4-Å resolution.";
RL Biochemistry 31:270-279(1992).
RN [15]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE: 90122799.
RA BYEON I.-J.L., KELLEY R.F., LINAS M.;
RT "1H NMR structural characterization of a recombinant kringle 2 domain
from human tissue-type plasminogen activator.";
RL Biochemistry 28:9350-9360(1989).
RN [16]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE: 91200042.
RA BYEON I.-J.L., KELLEY R.F., LINAS M.;
RT "Kringles-2 domain of the tissue-type plasminogen activator. 1H-NMR
assignments and secondary structure.";
RL Eur. J. Biochem. 197:155-165(1991).
RN [17]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE: 92106329.
RA BYEON I.-J.L., LINAS M.;
RT "Solution structure of the tissue-type plasminogen activator kringle
2 domain complexed to 6-aminohexanoic acid an antifibrinolytic
drug.";
RL J. Mol. Biol. 222:1035-1051(1991).
RN [18]
RP STRUCTURE BY NMR OF 38-85.
RX MEDLINE: 92292163.
RA DOWNING A.K., DRISCOLL P.C., HARVEY T.S., DUDGEON T.J., SMITH B.O.,
RA BARON M., CAMPBELL I.D.;
RT "Solution structure of the fibrin binding finger domain of
tissue-type plasminogen activator determined by 1H nuclear magnetic
resonance.";
RL J. Mol. Biol. 225:821-833(1992).
RN [19]
RP STRUCTURE BY NMR OF 36-126.
RX MEDLINE: 96027104.
RA SMITH B.O., DOWNING A.K., DRISCOLL P.C., DUDGEON T.J., CAMPBELL I.D.;
RT "The solution structure and backbone dynamics of the fibronectin type
I and epidermal growth factor-like pair of modules of tissue-type
plasminogen activator.";
RL Structure 3:823-833(1995).
RN [20]
RP SINGLE R-Y BOND CONVERTS PLASMINOGEN TO PLASMIN BY HYDROLYZING A
DESTRUCTION, PARTICULARLY IN FIBRINOLYSIS, AND IN CELL MIGRATION.
CC -I- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-1-VAL BOND IN
PLASMINOGEN TO FORM PLASMIN.
CC -I- SUBUNIT: CLEAVAGE AFTER ARG-310 BY PLASMIN OR TRYPSIN RESULTS IN
A TWO-CHAIN FORM OF THE MOLECULE. THESE TWO HALVES ARE HELD
TOGETHER BY ONE OR MORE DISULFIDE BONDS.
CC -I- ALTERNATIVE PRODUCTS: THERE EXISTS A SHORT VARIANT PROBABLY DUE TO
ALTERNATIVE SPLICING OF THE SAME GENE.
CC -I- MISCELLANEOUS: TPA ATTACHES TO THE KRINGLE STRUCTURE OF THE A

RESULT 7
ID PLMN_BOVIN STANDARD; PRT; 812 AA.
AC P06868; Q28162;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7).
GN PLG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN (1)
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC BERGLUND L., ANDERSEN M.D., PETERSEN T.E.;
RC Cloning and characterization of the bovine plasminogen cDNA.";
RC Int. Dairy J. 5:593-603(1995).
[2]
RP SEQUENCE OF 27-812, AND CARBOHYDRATE-BINDING SITES.
RX MEDLINE; 85203906.
RA SCHALLER J., MOSER P.W., DANNEGER-MULLER G.A.K., ROSSELET S.J.,
RA KAMFER U., RICKLI E.E.;
RA "Complete amino acid sequence of bovine plasminogen. Comparison with
RI human plasminogen.";
RI Eur. J. Biochem. 149:267-278(1985).
RN [3]
RP SEQUENCE OF 706-812 FROM N.A.
RX MEDLINE; 85023311.
RA MALINOWSKI D.P., SADLER J.E., DAVIE E.W.;
RA "Characterization of a complementary deoxyribonucleic acid coding for
RI human and bovine plasminogen.";
RI Biochemistry 23:4243-4250(1984).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION: IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79402; CAA55939.1; -
DR EMBL; K02935; AAA30714.1; -
DR PIR; A25835; PLBO.
DR HSSP; P00747; 2PK4.
DR PROSITE; PS00021; KRINGLE_1; 5.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PROSITE; PS00700; KRINGLE_2; 5.
DR PFAM; PF00051; kringle; 5.
DR PFAM; PF00089; trypsin; 1.
DR Hydrolase; Serine protease; Plasma; Glycoprotein; Fibrinolysis;
KW Tissue remodeling; Blood coagulation; Kringle; Zymogen; Signal.
FT SIGNAL 1 26
FT CHAIN 27 812 PLASMINOGEN.
FT CHAIN 27 583 HEAVY CHAIN A.

FT CHAIN 584 812 LIGHT CHAIN B.
FT DOMAIN 584 812 CATALYTIC.
FT DOMAIN 110 188 KRINGLE 1.
FT DOMAIN 192 269 KRINGLE 2.
FT DOMAIN 282 359 KRINGLE 3.
FT DOMAIN 384 461 KRINGLE 4.
FT DOMAIN 485 564 KRINGLE 5.
FT CARBOHYD 315 315
FT CARBOHYD 365 365
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM.
FT ACT_SITE 667 667 CHARGE RELAY SYSTEM.
FT ACT_SITE 762 762 CHARGE RELAY SYSTEM.
FT CONFLICT 335 335 N -> D (IN REF. 2).
FT CONFLICT 516 516 Q -> H (IN REF. 2).
FT CONFLICT 555 555 P -> L (IN REF. 2).
FT CONFLICT 744 744 T -> R (IN REF. 3).
SQ SEQUENCE 812 AA; 91216 MW; 383C0328 CRC32;

Query Match 28.5%; Score 174; DB 1; Length 812;
Best Local Similarity 35.9%; Pred. No. 1,05e-18;
Matches 23; Conservative 11; Mismatches 27; Indels 3; Gaps 3;

Db 294 VAYTESGTCQMRSEOTPKHNRTPEFPC-KNLENYGRNPGKAPMC-YT-TNSEVR 350
Qy 47 VSVTDFGAPCLMAEVPFLERSPPASWALRGQRHNFCRSPDGAGRPWCYGDAGKVD 106
Db 351 WEXC 354
Qy 107 WGYC 110

RESULT 8
ID PLMN_MOUSE STANDARD; PRT; 812 AA.
AC P20918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) [CONTAINS: ANGIOSTATIN].
GN PLG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 91184812.
RA DEGEN S.J., BELL S.M., SCHAEFER L.A., ELLIOTT R.W.;
RA "Characterization of the cDNA coding for mouse plasminogen and
RI localization of the gene to mouse chromosome 17.";
RI Genomics 8:49-61(1990).
RN [2]
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE; 95042728.
RA O'REILLY M.S., HOLMGREN L., SHING Y., CHEN C., ROSENTHAL R.A.,
RA MOSS W., LANE W.S., CAO Y., SAGE E.H., FOLKMAN J.;
RA "Angiostatin: a novel angiogenesis inhibitor that mediates the
RT suppression of metastases by a Lewis lung carcinoma.";
RT Cell 79:315-328(1994).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION: IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
CC NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
CC METASTATIC TUMORS IN VIVO.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.

```

Db      287  VSVVSKCTQKQSESDPHRNHRTPEKFC-KNLBENTCRNPDETAPEKPYTDTDSQLR-- 343
Oy      47  VSTYDFACPLKRAAEVPPFLERSPPASWAQLRGCRHNHCPSDAGRPKCFYGDARKVD 106
Db      344  WEYCE 348
Oy      107  WGYCD 111

RESULT          9
ID FA12_CAVPO STANDARD: PRT. 603 AA.
AC 004962;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COAGULATION FACTOR XII PRECURSOR (EC 3.4.21.36) (HAGEMAN FACTOR)
DE (NAF) (FRAGMENT).
GN F12.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriognathii; Caviidae; Cavia.
RN
RP TISSUE=LIVER;
RC SEQUENCE FROM N.A., AND SEQUENCE OF 19-37, 318-332 AND 359-373.
RX MEDLINE; 93003367.
RA SEMBA U., YAMAMOTO T., KUNISADA T., SHIBUYA Y., TANASE S.,
RA KAMBARA T., OKABE H.;
RT "Primary structure of guinea-pig Hageman factor: sequence around the
RT cleavage site differs from the human molecule.";
RL Biochim. Biophys. Acta 1159:113-121(1992).
CC -1- FUNCTION: FACTOR XII IS A SERUM GLYCOPROTEIN THAT PARTICIPATES IN
CC THE INITIATION OF BLOOD COAGULATION, FIBRINOLYSIS, AND THE
CC GENERATION OF BRADYKININ AND ANGIOTENSIN.
CC -1- CATALYTIC ACTIVITY: CLEAVES SELECTIVELY ARG-I-LIKE BONDS AND
CC ACTIVATES COAGULATION FACTORS VII AND XI.
CC -1- MISCELLANEOUS: FACTOR XII, PREKALLIKREIN, AND HMW KININOGEN FORM A
CC COMPLEX BOUND TO AN ANIONIC SURFACE. PREKALLIKREIN IS CLEAVED BY
CC FACTOR XII TO FORM KALLIKREIN, WHICH THEN CLEAVES FACTOR XII FIRST
CC TO ALPHA-FACTOR XIIA AND THEN TO BETA-FACTOR XIIA. ALPHA-FACTOR
CC XIIA ACTIVATES FACTOR XI TO FACTOR XIA.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE REGION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X68615; CAA48600.1; -.
CC DR HSSP; P00763; IDPO.
CC DR PROSITE; PS00021; KRINGLE_1; 1.
CC DR PROSITE; PS00022; EGF_1; 2.
CC DR PROSITE; PS00023; FIBRONECTIN_2; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC DR PROSITE; PS01186; EGF_2; 1.
CC DR PROSITE; PS01253; FIBRONECTIN_1; 1.
CC DR PROSITE; PS50070; KRINGLE_2; 1.
CC DR PRAM; PF00008; EGF_2; 1.
CC DR PRAM; PF00039; fn1; 1.
CC DR PRAM; PF00040; fn2; 1.
CC DR PRAM; PF00051; kringle; 1.
CC DR PRAM; PF00089; trypsin; 1.
CC KM Glycoprotein; Blood coagulation; Plasma; Kringle; Serine protease;
CC Hydrolyase; Fibrinolysis; EGF-like domain; Repeat; Zymogen; Signal.

```



```
FT NON_TER 1 18
FT SIGNAL <1 18
FT CHAIN 19 358
FT CHAIN 399 603
FT DOMAIN 46 87
FT DOMAIN 93 130
FT DOMAIN 132 172
FT DOMAIN 173 209
FT DOMAIN 216 294
FT DOMAIN 312 342
FT DOMAIN 359 603
FT ACT_SITE 398 398
FT ACT_SITE 447 447
FT ACT_SITE 551 551
FT ACT_SITE 551 551
FT DISULFID 97 109
FT DISULFID 103 118
FT DISULFID 120 129
FT DISULFID 134 162
FT DISULFID 160 169
FT DISULFID 177 188
FT DISULFID 182 197
FT DISULFID 199 208
FT DISULFID 216 294
FT DISULFID 237 276
FT DISULFID 265 289
FT DISULFID 345 472
FT DISULFID 383 399
FT DISULFID 391 461
FT DISULFID 422 425
FT DISULFID 488 557
FT DISULFID 520 536
FT DISULFID 547 578
FT CARBOHYD 248 248
FT CARBOHYD 270 270
FT CARBOHYD 419 419
SQ SEQUENCE 603 AA: 66795 MW: 35C16971 CRC32;
```

```
Query Match 28.0%; Score 171; DB 1; Length 603;
Best Local Similarity 37.3%; Pred. No. 4.57e-18;
Matches 22; Conservative 12; Mismatches 23; Indels 2; Gaps 2;
```

```
Db 234 GAKCORMASEATYRNMTAQLRGLGH-HTRCRNPNDNTRPWCYFWMG-NRLSWEYCD 230
QY 53 GACPLMAEVPFLERSPPASMAQLRGQRHNFRCSPDGAGRPWCFCFGDARGVDMGYCD 111
```

```
FT 10
PLMN_PTIG STANDARD: PRT: 790 AA.
P06867:
01-JAN-1988 (Rel. 06, Created)
01-FEB-1991 (Rel. 17, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
DE PLASMINOGEN (EC 3.4.21.7).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
NM Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1]
RA SEQUENCE OF 1-560.
RA SCHALLER J., MARTI T., ROSSSELET S.J., KAEMPFER U., RICKLI E.E.;
RT "Amino acid sequence of the heavy chain of porcine plasmin. Comparison
RT of the carboxydrate attachment sites with the human and bovine
RT species."
RL Fibrinolysis 1:91-102(1987).
[2]
RP SEQUENCE OF 450-790.
RX MEDLINE: 85203907.
RA MARTI T., SCHALLER J., RICKLI E.E.;
RT "Determination of the complete amino-acid sequence of porcine
RT miniplasminogen."
RL Eur. J. Biochem. 149:279-285(1985).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION.
```

```
CC AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAAFIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC PIR: A25834; A25834.
DR PIR: S03733; S03733.
DR HSSP: P00747; SHPG.
DR PROSITE: PS00134; TRYPSIN_HIS: FALSE_NEG.
DR PROSITE: PS00021; KRINGLE_1; 5.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS50070; KRINGLE_2; 5.
DR PFAM: PF00051; kringle; 5.
DR PFAM: PF00089; trypsin; 1.
KW Hydrolase. Serine protease. Plasma. Glycoprotein. Fibrinolysis;
KW Tissue remodeling; Blood coagulation; Kringle; Zymogen.
FT CHAIN 1 560
FT CHAIN 561 790
FT DOMAIN 561 790
FT DOMAIN 84 162
FT DOMAIN 166 243
FT DOMAIN 256 333
FT DOMAIN 358 435
FT DOMAIN 461 540
FT ACT_SITE 602 602
FT ACT_SITE 645 645
FT ACT_SITE 740 740
FT CARBOHYD 249 249
SQ SEQUENCE 790 AA: 88592 MW: EE597814 CRC32;
```

```
Query Match 28.0%; Score 171; DB 1; Length 790;
Best Local Similarity 37.5%; Pred. No. 4.57e-18;
Matches 24; Conservative 8; Mismatches 29; Indels 3; Gaps 3;
```

```
Db 268 VSVTASGTCQRMASQSPFKHNRTPEFPC-KNLENYCRNPDEGTAAPWC-YTTD-SEVR 324
QY 47 VSVTDEGAPCLMAEVPFLERSPPASMAQLRGQRHNFRCSPDGAGRPWCFCFGDARGVDMGYCD 106
```

```
Db 325 WDYC 328
QY 107 WGYC 110
```

```
RESULT 11
ID FA12_BOVIN STANDARD: PRT: 593 AA.
AC P98140:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COAGULATION FACTOR XII PRECURSOR (EC 3.4.21.38) (HAGEMAN FACTOR)
DE (HAF) (FRAGMENT).
GN F12.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae;
OC Bovinae; Bos.
[1]
RP SEQUENCE FROM N.A.
RX TISSUE-LIVER;
RX MEDLINE: 94242782.
RA SHIBUYA Y., SEMBA U., OKABE H., KAMBARA T., YAMAMOTO T.;
RT "Primary structure of bovine Hageman factor (blood coagulation factor
RT XII): comparison with human and guinea pig molecules."
RL Biochim. Biophys. Acta 1206:63-70(1994).
[2]
```


FT VARSPLIC 163 167 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 728 AA: 82944 MM: 967CDFIE CRC32:

Query Match 26.4%; Score 161; DB 1; Length 728;
 Best local Similarity 36.1%; Pred. No. 5,866-16;
 Matches 22; Conservative 10; Mismatches 26; Indels 3; Gaps 3;

Db 227 TEGKTCQRMDOQTPHRHKFLPERYPD-KGFDDNYCRNPDGKPRFWCYTLDP-DT-PWEY 283
 QY 50 TDFGAPCLRWAEVPPFLERSPPASMAQLRCGRHNFCSRSDGAGRPWCFFYGDARGKVDWGY 109
 Db 284 C 284
 QY 110 C 110

Search completed: Mon Mar 13 10:21:12 2000
 Job time : 7 secs.

 W O R L D

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Run on: Mon Mar 13 10:21:30 2000; Maspar time 14.19 Seconds
 356,626 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-147-947-6
 Description: (40-112) from US09147947A.pep (2 of 6)
 Perfect Score: 611
 Sequence: 1 CPAGEPWSVTDFGAPCLRM.....RPMCFYGDARGKVDWGYCDC 73

Scoring table: PAM 150
 Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: splrembl12
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
 9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 37.029; Variance 62.983; scale 0.568

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	207	33.9	946	13	Q07153	RECEPTOR TYROSINE KINASE	3.09e-24
2	188	30.8	597	11	Q07153	RECEPTOR TYROSINE KINASE	3.09e-24
3	173	28.3	716	11	P70521	HEPATOCTYTE GROWTH FACT	3.03e-20
4	167	27.3	560	4	Q14520	HGF ACTIVATOR LIKE PRO	3.72e-17
5	165	27.0	616	6	Q07507	FXII	6.13e-16
6	163	26.7	290	4	Q02935	HEPATOCTYTE GROWTH FACT	1.55e-15
7	163	26.7	296	4	Q14519	HEPATOCTYTE GROWTH FACT	3.92e-15
8	162	26.5	2869	6	Q13208	HEPATOCTYTE GROWTH FACT	3.92e-15
9	161	26.4	567	4	Q13208	HEPATOCTYTE GROWTH FACT	6.22e-15
10	161	26.4	711	11	Q14870	MACROPHAGE-STIMULATING	9.86e-15
11	161	26.4	728	11	Q04007	HEPATOCTYTE GROWTH FACT	9.86e-15
12	161	26.4	748	11	Q01652	HEPATOCTYTE GROWTH FACT	9.86e-15
13	160	26.2	806	6	Q18783	PLASMINOGEN	9.86e-15
14	159	26.0	710	13	Q09102	HEPATOCTYTE GROWTH FACT	1.56e-14
15	152	24.9	454	6	Q46506	APOLIPROTEIN A (FRAG	2.47e-14
16	150	24.5	717	13	P70006	HEPATOCTYTE GROWTH FACT	6.02e-13
17	149	24.4	215	13	Q42341	HGF ALPHA-CHAIN (FRAG	1.49e-12
18	149	24.4	726	13	Q09078	HEPATOCTYTE GROWTH FACT	2.34e-12
19	147	24.1	704	13	Q09085	HEPATOCTYTE GROWTH FACT	5.74e-12
20	147	24.1	810	4	Q15146	PLASMINOGEN PRECURSOR	5.74e-12

21	141	23.1	716	13	Q01691	GROWTH FACTOR LIVERIN	8.33e-11
22	140	22.9	132	4	Q16609	(APOARGC).	1.30e-10
23	138	22.6	334	6	Q46507	PLASMINOGEN (FRAGMENT)	3.13e-10
24	138	22.6	411	4	Q15844	UROKINASE-TYPE PLASMIN	3.13e-10
25	138	22.6	431	4	Q16618	UROKINASE PRECURSOR (E	3.13e-10
26	135	22.1	714	5	Q02001	NEUROSPECIFIC RECEPTOR	1.17e-09
27	135	22.1	724	5	Q06391	RECEPTOR TYROSINE KINASE	1.17e-09
28	128	20.9	202	13	Q09075	RECEPTOR TYROSINE KINASE	2.41e-08
29	127	20.8	944	11	Q02138	TISSUE-TYPE PLASMINOGEN	3.70e-08
30	124	20.3	210	4	Q13494	HGF AGONIST/ANTAGONIST	1.32e-07
31	123	20.1	214	6	Q09170	UROKINASE-TYPE PLASMIN	2.02e-07
32	122	20.0	211	11	Q05027	HEPATOCTYTE GROWTH FACT	3.08e-07
33	119	19.5	145	6	Q28911	APOLIPROTEIN A (FRAG	1.08e-06
34	116	19.0	685	5	Q24488	PROTHROMBIN PRECURSOR	3.76e-06
35	116	19.0	806	5	Q17576	NEUROTHROMBIN RECEPTOR	3.76e-06
36	116	19.0	806	5	Q17576	KIN-8 PROTEIN.	3.76e-06
37	116	19.0	937	11	Q02139	ROR1	3.76e-06
38	116	19.0	943	4	Q01974	PROTEIN-TYROSINE KINASE	3.76e-06
39	115	18.8	937	4	Q01973	PROTEIN-TYROSINE KINASE	5.67e-06
40	111	18.2	263	4	Q00318	PUTATIVE PROTEIN.	2.90e-05
41	111	18.2	385	5	Q25101	SERINE PROTEINASE.	2.90e-05
42	93	15.2	389	6	Q07887	REISSNER'S FIBER GLYCO	3.16e-02
43	93	15.2	1637	6	Q09XSV8	SCO-SPODIN (FRAGMENT)	3.16e-02
44	91	14.9	420	13	Q090504	THROMBIN.	6.81e-02
45	90	14.7	607	13	Q01001	THROMBIN.	5.52e-02

ALIGNMENTS

RESULT	ID	Query Match	Length	DB	ID	Description	Pred. No.
1	Q07153	33.9%	946	13	Q07153	RECEPTOR TYROSINE KINASE	3.09e-24
2	Q07153	30.8%	597	11	Q07153	RECEPTOR TYROSINE KINASE	3.09e-24
3	Q07153	28.3%	716	11	P70521	HEPATOCTYTE GROWTH FACT	3.03e-20
4	Q07153	27.3%	560	4	Q14520	HGF ACTIVATOR LIKE PRO	3.72e-17
5	Q07153	27.0%	616	6	Q07507	FXII	6.13e-16
6	Q07153	26.7%	290	4	Q02935	HEPATOCTYTE GROWTH FACT	1.55e-15
7	Q07153	26.7%	296	4	Q14519	HEPATOCTYTE GROWTH FACT	3.92e-15
8	Q07153	26.5%	2869	6	Q13208	HEPATOCTYTE GROWTH FACT	3.92e-15
9	Q07153	26.4%	567	4	Q13208	HEPATOCTYTE GROWTH FACT	6.22e-15
10	Q07153	26.4%	711	11	Q14870	MACROPHAGE-STIMULATING	9.86e-15
11	Q07153	26.4%	728	11	Q04007	HEPATOCTYTE GROWTH FACT	9.86e-15
12	Q07153	26.4%	748	11	Q01652	HEPATOCTYTE GROWTH FACT	9.86e-15
13	Q07153	26.2%	806	6	Q18783	PLASMINOGEN	9.86e-15
14	Q07153	26.0%	710	13	Q09102	HEPATOCTYTE GROWTH FACT	1.56e-14
15	Q07153	24.9%	454	6	Q46506	APOLIPROTEIN A (FRAG	2.47e-14
16	Q07153	24.5%	717	13	P70006	HEPATOCTYTE GROWTH FACT	6.02e-13
17	Q07153	24.4%	215	13	Q42341	HGF ALPHA-CHAIN (FRAG	1.49e-12
18	Q07153	24.4%	726	13	Q09078	HEPATOCTYTE GROWTH FACT	2.34e-12
19	Q07153	24.1%	704	13	Q09085	HEPATOCTYTE GROWTH FACT	5.74e-12
20	Q07153	24.1%	810	4	Q15146	PLASMINOGEN PRECURSOR	5.74e-12

Result	ID	Query Match	Length	DB	ID	Description	Pred. No.
1	Q035727	33.9%	946	13	Q035727	RECEPTOR TYROSINE KINASE	3.09e-24
2	Q035727	30.8%	597	11	Q035727	RECEPTOR TYROSINE KINASE	3.09e-24
3	Q035727	28.3%	716	11	P70521	HEPATOCTYTE GROWTH FACT	3.03e-20
4	Q035727	27.3%	560	4	Q14520	HGF ACTIVATOR LIKE PRO	3.72e-17
5	Q035727	27.0%	616	6	Q07507	FXII	6.13e-16
6	Q035727	26.7%	290	4	Q02935	HEPATOCTYTE GROWTH FACT	1.55e-15
7	Q035727	26.7%	296	4	Q14519	HEPATOCTYTE GROWTH FACT	3.92e-15
8	Q035727	26.5%	2869	6	Q13208	HEPATOCTYTE GROWTH FACT	3.92e-15
9	Q035727	26.4%	567	4	Q13208	HEPATOCTYTE GROWTH FACT	6.22e-15
10	Q035727	26.4%	711	11	Q14870	MACROPHAGE-STIMULATING	9.86e-15
11	Q035727	26.4%	728	11	Q04007	HEPATOCTYTE GROWTH FACT	9.86e-15
12	Q035727	26.4%	748	11	Q01652	HEPATOCTYTE GROWTH FACT	9.86e-15
13	Q035727	26.2%	806	6	Q18783	PLASMINOGEN	9.86e-15
14	Q035727	26.0%	710	13	Q09102	HEPATOCTYTE GROWTH FACT	1.56e-14
15	Q035727	24.9%	454	6	Q46506	APOLIPROTEIN A (FRAG	2.47e-14
16	Q035727	24.5%	717	13	P70006	HEPATOCTYTE GROWTH FACT	6.02e-13
17	Q035727	24.4%	215	13	Q42341	HGF ALPHA-CHAIN (FRAG	1.49e-12
18	Q035727	24.4%	726	13	Q09078	HEPATOCTYTE GROWTH FACT	2.34e-12
19	Q035727	24.1%	704	13	Q09085	HEPATOCTYTE GROWTH FACT	5.74e-12
20	Q035727	24.1%	810	4	Q15146	PLASMINOGEN PRECURSOR	5.74e-12

As hit 1, maybe

```

AC      035727.7-   (Tremblrel. 05. Created)
DT      01-JAN-1998
DT      01-FEB-1997 (Tremblrel. 02. Last sequence update)
DT      01-JAN-1998 (Tremblrel. 05. Last sequence update)
DT      01-NOV-1999 (Tremblrel. 12. Last annotation update)
DE      FACTOR XII.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RA      SCHLOSSER M., SCHWAGER S., ENGEL W.;
RL      Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
DR      EMBL: X99571; CA67891.1; -.
DR      HSSP: P00760; IAO7.
DR      PROSITE: PS00023; FIBRONECTIN_2; 1.
DR      PROSITE: PS01253; FIBRONECTIN_1; 1.
DR      PFAM: PF00040; fn2; 1.
DR      PFAM: PF00089; trypsin; 1.
DR      PFAM: PF00051; kringle; 1.
DR      PFAM: PF00008; EGF; 2.
DR      PFAM: PF00039; fn1; 1.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
DR      PRINTS: PR00018; KRINGLE.
SQ      SEQUENCE 597 AA; 65638 MW; 1EA8CD44 CRC32;

Query Match          30.8%; Score 188; DB 11; Length 597;
Best Local Similarity 40.0%; Pred. No. 3,03e-20;
Matches 26; Conservative 13; Mismatches 20; Indels 6; Gaps 6

Db      230 GTTGAGCACCRTVATYATRNMTKALSMG-L-GH-HAFCRPNDDTYPCFWSMG-DRL 285
Qy      48 SVTDGACGLRAEAVPEFLERSP-PA-SWAOLRGGRHNHCSPDAGRWCFEGDARGKV 105
        ::::|||||::: :::: |::: |::| ||||| ||||| :::
Db      286 SWDYC 290
Qy      106 DWGYC 110

RESULT 3
ID      P70521 PRELIMINARY; PRT; 716 AA.
DC      P70521.
DT      01-FEB-1997 (Tremblrel. 02. Created)
DT      01-FEB-1997 (Tremblrel. 02. Last sequence update)
DT      01-NOV-1999 (Tremblrel. 12. Last annotation update)
DE      HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN PRECURSOR
DE      (MACROPHAGE STIMULATORY PROTEIN) (MSP).
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RA      OHSHIRO K., IWAMA A., MATSUNO K., EZAKI T., SAKAMOTO O., HAMAGUCHI I.,
RA      KASAHARA N., SUDA T.;
RL      Molecular cloning of rat macrophage-stimulating protein and its
RL      involvement in the male reproductive system."
RL      Biochem. Biophys. Res. Commun. 227:273-280(1996).
CC      -!- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC      CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
CC      CONSERVED.
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
DR      EMBL: X95096; CA64473.1; -.
DR      HSSP: P00747; IPMK.
DR      PFAM: PF00051; kringle; 4.
DR      PFAM: PF00089; trypsin; 1.
DR      PRINTS: PR00018; KRINGLE.
DR      PRINTS: PR00722; CHYMOTRYPSIN.
KW      Signal.
TT      SIGNAL 1 31 POTENTIAL.
```

```

SQ CHAIN 32 716 MACHOPAGE STIMULATING PROTEIN.
ET SEQUENCE 716 AA; 80733 MW; 7C8133F6 CRC32;

Query Match
Best Local Similarity 28.3%; Score 173; DB 11; Length 716;
Matches 22; Conservative 16; Mismatches 24; Indels 3; Gaps 2

Db 203 VDVTSSGECRCRMDLQHSHSPHFPEKPD-KALKDNCRRPDASERPWCTTDP--NVE 259
OY 47 VSVDFGAPCLRMAVEPPLERSPPASWAQLRGGRHNFRCSPDAGRWCFYGDARGKVD 106
Db 260 REFC 264
OY 107 WCYCD 111

RESULT 4
ID 014520 PRELIMINARY; PRT; 560 AA.
AC 014520; 000663;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HGF ACTIVATOR LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA CHOI-MIURA N.H.; TOBE T., SUMIYA J., NAKANO Y., SANO Y., MAZDA T.,
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96425001.
RA CHOI-MIURA N.H.; TOBE T., SUMIYA J., NAKANO Y., SANO Y., MAZDA T.,
RA TOMITA M.;
RT "Purification and characterization of a novel hyaluronan-binding
RT protein (PHB) from human plasma. It has three EGF, a kringle and a
RT serine protease domain, similar to hepatocyte growth factor
RT activator.";
RL J. Biochem. 119:1157-1165(1996).
DR EMBL; D49742; BAA08576.1; -.
DR HSSP; P00763; IDPO.
DR HSSP; P00763; IDPO.
DR PFAM; PF00008; EGF; 3.
DR PFAM; PF00051; kringle; 1.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00018; KRINGLE.
DR PRINTS; PR00722; CHYMOTRYPSIN.
SO SEQUENCE 560 AA; 62671 MW; 4AC81907 CRC32;

Query Match
Best Local Similarity 27.3%; Score 167; DB 4; Length 560;
Matches 17; Conservative 4; Mismatches 9; Indels 0; Gaps 0.

Db 243 HNECRNDPDADEKPMCFIKVINDKVKWEXCD 272
OY 82 HNECRSPDAGRPWCYFGDARGKVDWGYCD 111

RESULT 5
ID 097507 PRELIMINARY; PRT; 616 AA.
AC 097507;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE Fx11.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA TAKAHASHI T., KIHARA T.;

```


RT "Porcine liver factor XII."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB022426; BAA37148.1; -
 DR HSSP: P00763; IDPO.
 DR PROSITE: PS00023; FIBRONECTIN_2; 1.
 DR PROSITE: PS01253; FIBRONECTIN_1; 1.
 SQ SEQUENCE 616 AA: 68012 MW: 31CCD856 CRC32;

Query Match 27.0%; Score 165; DB 6; Length 616;
 Best Local Similarity 40.0%; Pred. No. 1,55e-15;
 Matches 24; Conservative 11; Mismatches 19; Indels 6; Gaps 6;

DB 235 GACPQWASATYWNMTAEOALWG-L-GD-HAFCRNPNDTRPCFWRG-DOLSWQYC 290
 53 GADPLRMAEVPFLERSP-PA-SWAQLRGQRHNFRCSPDAGRPWCYGDARGKVDWGYC 110

AC 002935 PRELIMINARY; PRT; 290 AA.
 AC 002935;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
 DE HEPATOCYTE GROWTH FACTOR, HEAVY CHAIN PRECURSOR.
 GN HGF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RA MEDLINE: 91200041.
 RA MIYAZAWA K., KITAMURA A., NAKA D., KITAMURA N.;
 RT "An alternatively processed mRNA generated from human hepatocyte
 growth factor gene";
 RL Eur. J. Biochem. 197;15-22(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93087571.
 RA HARTMANN G., NALDINI L., WEIDNER K.M., SACHS M., VIGNA E.,
 RA COMOLLO P.M., BIRDMETER W.;
 RT "A functional domain in the heavy chain of scatter factor/hepatocyte
 growth factor binds the c-Met receptor and induces cell dissociation
 but not mitogenesis";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11574-11578(1992).
 CC -1- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
 HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPIC FACTOR, AND ACTS AS
 GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES. IT
 HAS NO DETECTABLE PROTEASE ACTIVITY.
 -1- SIMILARITY: CONTAINS TWO KRINGLE REGIONS.
 -1- SIMILARITY: HIGH, TO OTHER HGF, LOWER, TO PLASMINOGEN.

CC EMBL: X57574; CAA40802.1; -
 DR EMBL: 102931; AAA52649.1; -
 DR HSSP: P14210; 1BHT.
 DR PFAM: PF00051; kring1e; 2.
 DR PRINTS: PR00018; KRINGLE.
 KW Growth factor; Kring1e; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 290
 FT DOMAIN 128 206
 FT DOMAIN 211 288
 FT MOD_RES 32 32
 FT FT
 SO SEQUENCE 290 AA: 33765 MW: 9EF113DE CRC32;

Query Match 26.7%; Score 163; DB 4; Length 290;
 Best Local Similarity 36.1%; Pred. No. 3,92e-15;
 Matches 22; Conservative 11; Mismatches 25; Indels 3; Gaps 2;

DB 226 TESGKICQRMWDQTPRHKFLPERYPD-KGFDNYCRNPDGQRPWCYTLDPHTR--WEY 282
 50 TDFGAPCLRMAEVPFLERSPASMQLRGQRHNFRCSPDAGRPWCYGDARGKVDWGY 109

DB 283 C 283
 QY 110 C 110

RESULT 7
 ID 014519 PRELIMINARY; PRT; 296 AA.
 AC 014519;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
 DE COMPETITIVE HGF ANTAGONIST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CHAN A.M.L., ROBIN J.S., BOTTARO D.P., HIRSCHFELD D.W., CHEDID M.,
 RA AARONSON S.A.;
 RL Science 0:0-0(0).
 DR EMBL: M77227; AAA35980.1; -
 DR HSSP: P14210; 1BHT.
 DR PFAM: PF00051; kring1e; 2.
 DR PRINTS: PR00018; KRINGLE.
 SQ SEQUENCE 296 AA: 34546 MW: 202394A7 CRC32;

Query Match 26.7%; Score 163; DB 4; Length 296;
 Best Local Similarity 36.1%; Pred. No. 3,92e-15;
 Matches 22; Conservative 11; Mismatches 25; Indels 3; Gaps 2;

DB 226 TESGKICQRMWDQTPRHKFLPERYPD-KGFDNYCRNPDGQRPWCYTLDPHTR--WEY 282
 50 TDFGAPCLRMAEVPFLERSPASMQLRGQRHNFRCSPDAGRPWCYGDARGKVDWGY 109
 DB 283 C 283
 QY 110 C 110

RESULT 8
 ID 028398 PRELIMINARY; PRT; 2869 AA.
 AC 028398;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
 DE APOLOPROTEIN (FRAGMENT).
 OS Erinaceus europaeus (Western European hedgehog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Insectivora; Erinaceidae; Erinaceinae; Erinaceus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 96025778.
 RA LAWN R.M., BOONMARK N.W., SCHWARTZ K., LINDAHL G.E., WADE D.P.,
 RA BYRNE C.D., FONG K.J., MEER K., PATTHY L.;
 RT "The recurring evolution of lipoprotein(a). Insights from cloning of
 hedgehog apolipoprotein(a).";
 RL J. Biol. Chem. 270:24004-24009(1995).
 DR EMBL: U33170; AAC48522.1; -
 DR HSSP: P00747; 1PMK.
 DR PFAM: PF00051; kring1e; 31.
 KW Lipoprotein.
 FT NON_TER 1 1
 SO SEQUENCE 2869 AA: 318601 MW: FCC166B9 CRC32;

Query Match 26.5%; Score 162; DB 6; Length 2869;
 Best Local Similarity 35.9%; Pred. No. 6,22e-15;
 Matches 23; Conservative 9; Mismatches 29; Indels 3; Gaps 3;

DB 2417 MAYTASGHICQRMRESPSHSHTPENYPT-KLVYCNCRNPDGEVAPNC-YTTNSA-VR 2473
 47 VSVTDGAPCLRMAEVPFLERSPASMQLRGQRHNFRCSPDAGRPWCYGDARGKVD 106

T	14
D	.9140Z.
C	PRELIMINARY;
E	PRT; 710 AA.
F	
G	
H	
I	01-NOV-1996 (Tremblrel. 01, Created)
J	01-NOV-1996 (Tremblrel. 01, last sequence update)
K	DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
L	DE HEPATOCYTE GROWTH FACTOR (FRAGMENT).
M	OS Xenopus laevis (African clawed frog).
N	CC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Amphibia;
O	Batrachia; Anura; Mesobatrachia; Piploidea; Pipoidea; Xenopodinae;
P	Xenopus.
Q	
R	141
S	SEQUENCE FROM N.A.
T	NT MEDLINE: 95267690.
V	RA NAKAMURA H., TASHIRO K., NAKAMURA T., SHIKAWA K.;
W	"Molecular cloning of xenopus hgf cdna and its expression studies in xenopus early embryogenesis."
X	EMBL: S77422; AAB84354.1; .
Y	DR HSSP: P14210; IBHT.
Z	PRAM: PF00051; kringein; 4.
[blank]	PFAM: PF00089; trypsin; 1.
[blank]	NON_TER 710 710
[blank]	SEQUENCE 710 AA; 81487 MW; 791160ZF CRC32;

```

Matches 11; Conservative 10; Mismatches 27; Indels 3; Gaps 3
Db 213 TESGKEQKRDLDLPFHKKRPKEYPN-KGLNDNYCNPDGKSRPWCYTIDP-DT-SMEF 269
QY 50 TDFAPCLRMAEVPFLERSPSPASMAQLRGQRHNFCRSPDGAGRPMCFYGDAGKYDWG 109
Db 270 C 270
QY 110 C 110

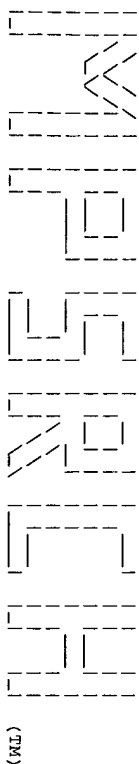
RESULT 15
ID 046506 PRELIMINARY; PRT: 454 AA.
AC 046506:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE APOLIPOPROTEIN A (FRAGMENT).
GN BABAPOA.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheciinae;
OC Papio.
OC [1]
RN RN
RP SEQUENCE FROM N.A.
RA COX I.A., JETT C., HIXSON J.E.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF029691; AAB97886.1; -.
DR HSSP; P00747; 2PK4.
DR PFAM; PF000051; kringl1; 2.
DR PFAM; PF000089; trypsin; 1.
DR Lipoprotein.
FT NON_TER 1 1
SQ SEQUENCE 454 AA; 50041 MW; 2400BE6C CRC32;

Query Match 24.9%; Score 152; DB 6; Length 454;
Best Local Similarity 33.9%; Pred. No. 6.02e-13;
Matches 20; Conservative 12; Mismatches 24; Indels 3; Gaps 3;
Db 6 GRNCQAWSSMTPHOHSRTPKNYRNA-GLTRNYCRNPDAEIRPWCYTWDP-S-YRWEXCN 61
QY 53 GAPCLRMAEVPFLERSPSPASMAQLRGQRHNFCRSPDGAGRPMCFYGDAGKYDWG 111

Search completed: Mon Mar 13 10:21:47 2000
Job time : 17 secs.

```

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-p protein - protein database search, using Smith-Waterman algorithm
Mon Mar 13 10:19:31 2000; MasPar time 8.11 Seconds
391.512 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (378-822) from US09147947A.pep (1 of 6)
Perfect Score: 1892
Sequence: 1 IIGGKNSLRGSGMPQVSLRL.....FGVYTKVSAFPMIKSVTKL 245

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PC9_COMB 4:backfile1

Statistics: Mean 30.953; Variance 131.070; scale 0.236

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	594	31.4	638	US-08-681- Sequence 3, Applicatio	4.46e-46
2	554	29.3	356	US-08-681- Sequence 1, Applicatio	3.65e-42
3	545	28.8	430	US-07-942- Sequence 3, Applicatio	2.76e-41
4	543	28.7	306	US-08-560- Sequence 45, Applicati	4.32e-41
5	543	28.7	331	US-08-560- Sequence 46, Applicati	4.32e-41
6	543	28.7	365	US-08-720- Sequence 83, Applicati	4.32e-41
7	543	28.7	365	US-08-093- Sequence 83, Applicati	4.32e-41
8	543	28.7	393	US-08-560- Sequence 44, Applicati	4.32e-41
9	543	28.7	411	US-08-087- Sequence 1, Applicati	4.32e-41
10	543	28.7	411	US-08-286- Sequence 18, Applicati	4.32e-41
11	543	28.7	411	US-08-153- Sequence 18, Applicati	4.32e-41
12	543	28.7	430	US-08-153- Sequence 18, Applicati	4.32e-41
13	543	28.7	431	US-08-153- Sequence 18, Applicati	4.32e-41
14	543	28.7	431	US-08-153- Sequence 18, Applicati	4.32e-41
15	543	28.7	431	US-08-153- Sequence 18, Applicati	4.32e-41
16	533	28.2	335	US-08-811- Sequence 47, Applicati	3.26e-40
17	533	28.2	335	US-08-811- Sequence 47, Applicati	3.26e-40
18	533	28.2	437	US-08-811- Sequence 57, Applicati	4.08e-40
19	533	28.2	527	US-08-811- Sequence 51, Applicati	4.08e-40
20	530	28.0	254	US-08-560- Sequence 39, Applicati	4.08e-40
21	530	28.0	411	US-08-560- Sequence 49, Applicati	8.00e-40
22	527	27.9	253	US-09-027- Sequence 48, Applicati	8.00e-40
23	527	27.9	354	US-08-811- Sequence 61, Applicati	1.57e-39

24	527	27.9	355	4	5223256-1	Patent No. 5223256.	1.57e-39
25	527	27.9	355	1	US-08-217-	Sequence 1, Applicatio	1.57e-39
26	527	27.9	355	1	US-08-427-	Sequence 6, Applicatio	1.57e-39
27	527	27.9	355	2	US-08-811-	Sequence 45, Applicati	1.57e-39
28	527	27.9	355	1	US-08-427-	Sequence 2, Applicatio	1.57e-39
29	527	27.9	355	1	US-08-137-	Sequence 1, Applicatio	1.57e-39
30	527	27.9	355	1	US-08-217-	Sequence 1, Applicatio	1.57e-39
31	527	27.9	356	1	US-08-427-	Sequence 8, Applicatio	1.57e-39
32	527	27.9	383	2	US-08-427-	Sequence 4, Applicatio	1.57e-39
33	527	27.9	389	2	US-08-558-	Sequence 6, Applicatio	1.57e-39
34	527	27.9	437	2	US-08-811-	Sequence 67, Applicati	1.57e-39
35	527	27.9	527	1	US-08-811-	Sequence 55, Applicati	1.57e-39
36	527	27.9	527	1	US-08-811-	Sequence 55, Applicati	1.57e-39
37	527	27.9	527	1	US-08-811-	Sequence 55, Applicati	1.57e-39
38	527	27.9	527	1	US-08-811-	Sequence 55, Applicati	1.57e-39
39	527	27.9	527	1	US-08-811-	Sequence 55, Applicati	1.57e-39
40	527	27.9	562	4	5244676-5	Patent No. 5244676.	1.57e-39
41	527	27.9	562	2	US-08-560-	Sequence 50, Applicati	1.57e-39
42	527	27.9	562	2	US-08-883-	Sequence 38, Applicati	1.57e-39
43	527	27.9	562	4	5185259-3	Patent No. 5185259.	1.57e-39
44	527	27.9	562	4	5200340-2	Patent No. 5200340.	1.57e-39
45	527	27.9	562	2	US-08-811-	Sequence 43, Applicati	1.57e-39

ALIGNMENTS

RESULT 1
ID US-08-681-151-3 STANDARD: PRT: 638 AA..
XX xxxxxx
AC
DT
XX
XX
DE Sequence 3, Application US/08681151
CC
CC Sequence 3, Application US/08681151
CC Patent No. 5189637
CC GENERAL INFORMATION:
CC APPLICANT: Au-Young, Janice
CC APPLICANT: Bandman, Olga
CC APPLICANT: Braxton, Scott Michael
CC APPLICANT: Goli, Surya
CC TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
CC NUMBER OF SEQUENCES: 4
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: US
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ Version 1.5
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/681,151
CC FILING DATE: Herewith
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0074US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:

```
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/942,157A
```


THIS PAGE BLANK (USPTO)


```

FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 875 ZMOGEN ACTIVATION REGION.
FT ACT SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT SITE 630 631 CHARGE RELAY SYSTEM.
FT ACT SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 6752272B CRC32:

Query Match
Best Local Similarity 100.0%; Score 1892; DB 1; Length 875;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 631 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSTAY 690
578 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSTAY 637
691 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRSDYDIALVRLQGPBEOCARFSSHYLPAC 750
638 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRSDYDIALVRLQGPBEOCARFSSHYLPAC 697
Db 751 LPLMRERPKTASNCYITGMDTGRAYSRTLQQAIPLLPKRCEERYKGRFTGRMLCAG 810
Qy 698 LPLMRERPKTASNCYITGMDTGRAYSRTLQQAIPLLPKRCEERYKGRFTGRMLCAG 757
Db 811 NHEHKRVDSGCGDGGPLMCEPESWVYGVTSWGYGCGVDSPEVYTKVSAFVPMIK 870
Qy 758 NHEHKRVDSGCGDGGPLMCEPESWVYGVTSWGYGCGVDSPEVYTKVSAFVPMIK 817
Db 871 SVTKL 875
Qy 818 SVTKL 822

RESULT 2
ID NCTR_MOUSE STANDARD: PRT: 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE NEURORPEPSIN PRECURSOR (EC 3.4.21.-) (MOTOPISIN) (BRAIN-SPECIFIC SERINE
DE PROTEASE 3) (BSP-3).
DE PRESS12.
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC [1]
SEQUENCE FROM N.A.
TISUE-BRAIN;
MEDLINE: 97401523.
RA GOSCHWEND T.P., KUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEBERGER P.;
RT "Neurocrystin, a novel multidomain serine protease expressed in the
RT nervous system."
RT Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT kringle-like structure and three scavenger receptor cysteine-rich
RT motifs."
RT Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
CC LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
CC AND AMYGALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```

```

CC CC
CC TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
-----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
CC EMBL: Y13192; CAA33646.1;
CC EMBL: D89871; BAA23986.1;
CC MGI: MGI:1100881; PRSS12.
CC PFAM: PF00530; SRCR; 3.
CC PFAM: PF00089; trypsin; 1.
CC DR PROSITE: PS00134; TRYPsin_HIS; 1.
CC DR PROSITE: PS00135; TRYPsin_SER; 1.
CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
CC KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; signal.
FT SIGNAL 1 21
FT CHAIN 1 21
FT DOMAIN 22 761
FT DOMAIN 85 157
FT DOMAIN 166 267
FT DOMAIN 273 373
FT DOMAIN 386 487
FT DOMAIN 505 761
FT DOMAIN 505 516
FT ACT_SITE 516 517
FT ACT_SITE 562 562
FT ACT_SITE 612 612
FT ACT_SITE 711 711
FT DISULFID 505 636
FT CARBOHYD 93 93
FT CARBOHYD 521 521
FT CARBOHYD 569 569
SQ SEQUENCE 761 AA: 84118 MW: 3F3CAF35 CRC32:

Query Match
Best Local Similarity 92.6%; Score 1752; DB 1; Length 761;
Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;

Db 517 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSTAY 576
Qy 578 IIGKNSLKGMPQWQSLRLKSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSRSTAY 637
Db 577 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRSDYDIALVRLQGPBEOCARLSTHYLPAC 636
Qy 638 RVGDYHTLVPEEFEEEDIGVQOIVIHREYRPRSDYDIALVRLQGPBEOCARFSSHYLPAC 697
Db 637 LPLMRERPKTASNCYITGMDTGRAYSRTLQQAIPLLPKRCEERYKGLFTGRMLCAG 696
Qy 698 LPLMRERPKTASNCYITGMDTGRAYSRTLQQAIPLLPKRCEERYKGRFTGRMLCAG 757
Db 697 NLEEDNRVDSGCGDGGPLMCEPESWVYGVTSWGYGCGVDSPEVYTKVSAFVPMIK 756
Qy 758 NLEHKRVDSGCGDGGPLMCEPESWVYGVTSWGYGCGVDSPEVYTKVSAFVPMIK 817
Db 757 SVTKL 761
Qy 818 SVTKL 822

RESULT 3
ID FALL_HUMAN STANDARD: PRT: 625 AA.
AC P03951;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DE 15-DEC-1999 (Rel. 39, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
DE ANTECEDENT) (PTA).
GN F11.

```


QY 753 MLCAGNLEHHRVDSGQSGPLMCCERGESWVYGVTSWGGCVKDSPGVYTVASAF 812
 Db 615 VDWI 618
 QY 813 VPWI 816

RESULT 4
 ID KAL_MOUSE STANDARD: PRT: 638 AA.
 AC P26262:
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KIK3 OR PK.
 OS Mus musculus (Mouse).
 RA Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RA Mammalia; Rodentia; Sciurognathi; Muridae; Mus.
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX STRAIN-BALB/C; TISSUE-LIVER:
 RA MEDLINE: 91090844.
 RA SEDDAH N.G., SAWYER N., HAMELIN J., MION P., BEAUBIEN G.,
 RA BRACHDAPA L., ROCHMONT J., MBIKAY M., CHRETIEN M.;
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species.";
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HWM KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HWM KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC MBL: M58588; AAA6393.1; -
 CC LIR: A36557; KOMSPL.
 CC HSSP: P00750; LRPE.
 CC MGD: MGI:102849; KIK3.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00495; APPLE; 4.
 DR PFAM; PF00024; apple; 4.
 DR PFAM; PF00089; trypsin; 1.
 KW Hydrolyase; Serine protease; Glycoprotein; Plasma; zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Duplication; Bradykinin.
 KM
 FT CHAIN 1 19
 FT CHAIN 20 390
 FT CHAIN 391 638
 FT REPEAT 20 105
 FT REPEAT 110 195
 FT REPEAT 200 285
 FT REPEAT 291 376
 FT DOMAIN 389 621
 FT CARBOHYD 127 127
 FT CARBOHYD 308 308
 FT CARBOHYD 396 396

PLASMA KALLIKREIN HEAVY CHAIN.
 PLASMA KALLIKREIN LIGHT CHAIN.
 APPLE 1.
 APPLE 2.
 APPLE 3.
 APPLE 4.
 CATALYTIC.
 PROBABLE.
 PROBABLE.
 PROBABLE.

FT CARBOHYD 453 453 PROBABLE.
 FT CARBOHYD 494 494 PROBABLE.
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
 FT DISULFID 21 104 BY SIMILARITY.
 FT DISULFID 47 77 BY SIMILARITY.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 363 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71368 MW; 8370CD2E CRC32;

Query Match 31.7%; Score 600; DB 1; Length 638;
 Best Local Similarity 40.6%; Pred. No. 3,41e-125;
 Matches 99; Conservative 54; Mismatches 73; Indels 18; Gaps 16;

Db 391 YVGTASTAGEMPWQVSLQVLY-SQTHL-CGSGITIGRWNTAHCADGIPYRWY 448
 QY 578 IIGKNSLGGMPWQVSLKSSHDGRLCGATLLSCWVTAACR--RYGNSTSY 635
 Db 449 GGILS-L-SEIKREPPSR-IRELLIHOEYKSEGNVIALIKILOTPLN-YTEF--OK-P 501
 QY 636 AVAVGYHILVBEFEELGVQOYIYHREYRBDYDALVLRDGPPEGCAFSHYLP 695
 Db 502 ICIP-SKADNTTYNCWVYGVYKKEGETONILOKATIPLVNPECQKRYDYINKQ 560
 QY 696 ACPLPERRPQRTASNCYITGMGT-GRAYST-LQQAIPPLPRFCEERYKGRFTGR- 752
 Db 561 MICAGY-KEGG-TDACKSGSPVCKHSG-RWLVGITSNMGEGGRDQGVYTKSEY 617
 QY 753 MLCAGNLEHHRVDSGQSGPLMCCERGESWVYGVTSWGGCVKDSPGVYTVASAF 812

Db 618 MDWI 621
 QY 813 VPWI 816

RESULT 5
 ID KAL_HUMAN STANDARD: PRT: 638 AA.
 AC P03952:
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KIK3.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OS Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86243359.
 RA CHUNG D.W., FUJIKAWA K., MCMULLEN B.A., DAVIE E.W.;
 RT "Human plasma prekallikrein, a zymogen to a serine protease that
 RT contains four tandem repeats.";
 RL Biochemistry 25:2410-2417(1986).
 RN (2)
 RP PARTIAL SEQUENCE. AND DISULFIDE BONDS.
 RX MEDLINE: 91152016.
 RA MCMULLEN B.A., FUJIKAWA K., DAVIE E.W.;

[illegible]

Oy	634	SVAR-VG-DVHTLVPSEFEEDIGVOQVIHREVRPDRSDYDIALVLRQSGEBCARRSS	691
Dd	156	-IGECPLQFQFAAPPRAPOTCWGTGWGLKKEGPRTSPTLOEARVALIDELCNSTRWTYN	214
Oy	692	HVLPAFLMEEREPDKTASNCYTITGWG---DTGRAYSRTLQOAAIPLPKRFCEE-R-YK	746
Dd	215	GRINSTNWCACAPRG-K-IDPCGGDSGGLMCRBRAENTFYVVGITISGVCAAKAPRGV	272
Oy	747	GRTFRMLCAENLNHHKKRVSDCGSGDGSLPMC-ERPESWVAVGTSTWGVCYGVKDSPGV	805
Dd	273	YTSTWPLYNNIAS	285
Oy	806	YTKSAFPWPWK	818
RESULT	8	UROT BOVIN STANDARD; PRT; 566 AA.	
ID	AC	028198:	
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	01-NOV-1997	(Rel. 35, Last annotation update)	
DE	TISSUE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.68) (TPA)		
DE	(T-PLASMINOGEN ACTIVATOR).		
GN	PLAT.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae;		
CC	Bovinae; Bos.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-KIDNEY;		
RA	RAVN P., BERGLUND L., PETERSEN T.E.;		
RT	"Cloning and characterization of the bovine plasminogen activators uPa		
RT	and tPa."		
RL	Ipt. Dairy J. 5:605-617(1995).		
CC	-1- FUNCTION: TPA CONVERTS PLASMINOGEN TO PLASMIN BY HYDROLYZING A		
CC	SINGLE R-V BOND IN PLASMINOGEN. ACTIVE IN TISSUE REMODELING AND		
CC	DESTRUCTION, PARTICULARLY IN FIBRINOLYSIS, AND IN CELL MIGRATION.		
CC	-1- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-1-VAL BOND IN		
CC	PLASMINOGEN TO FORM PLASMIN.		
CC	-1- MISCELLANEOUS: TPA ATTACHES TO THE KRINGLE STRUCTURE OF THE A		
CC	CHAIN OF FIBRIN.		
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.		
CC	-1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.		
CC	-1- SIMILARITY: CONTAINS 2 KRINGLE REGIONS.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRYPSIN FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: X85800; CAAS9795.1; -		
DR	HSP: P00750; IRTF.		
DR	PROSITE: PS00134; TRYPSIN_HIS_1.		
DR	PROSITE: PS00135; TRYPSIN_SER_1.		
DR	PROSITE: PS00022; EGF_1; 1.		
DR	PROSITE: PS01186; EGF_2; 1.		
DR	PROSITE: PS01253; FIBRONECTIN_1; 1.		
DR	PROSITE: PS00021; KRINGLE_1; 1.		
DR	PROSITE: PSS0070; KRINGLE_2; 2.		
DR	PFAM: PF00008; EGF_1.		
DR	PFAM: PF00039; fnl_1.		
DR	PFAM: PF00051; kringle_2.		
DR	PFAM: PF00089; trypsin_1.		
KW	Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;		
KW	plasma; kringle; EGF-like domain; Signal.		
FT	SIGNAL	1 21	
FT	PROPEL	22 33	
FT	BY SIMILARITY.		
FT	BT SIMILARITY.		

FT	CHAIN	34	314	TISSUE PLASMINOGEN ACTIVATOR A CHAIN (HEAVY).
FT	CHAIN	315	566	TISSUE PLASMINOGEN ACTIVATOR B CHAIN (LIGHT, SERINE PROTEASE).
FT	CHAIN			FIBRONECTIN TYPE-1.
FT	DOMAIN	40	82	EGF-LIKE.
FT	DOMAIN	83	121	KRINGLE 1.
FT	DOMAIN	128	209	KRINGLE 2.
FT	DOMAIN	219	300	
FT	ACT_SITE	361	361	CHARGE RELAY SYSTEM.
FT	ACT_SITE	410	410	CHARGE RELAY SYSTEM.
FT	ACT_SITE	517	517	CHARGE RELAY SYSTEM.
FT	DISULFID	42	72	BY SIMILARITY.
FT	DISULFID	70	79	BY SIMILARITY.
FT	DISULFID	87	98	BY SIMILARITY.
FT	DISULFID	92	109	BY SIMILARITY.
FT	DISULFID	111	120	BY SIMILARITY.
FT	DISULFID	128	209	BY SIMILARITY.
FT	DISULFID	149	191	BY SIMILARITY.
FT	DISULFID	180	204	BY SIMILARITY.
FT	DISULFID	219	300	BY SIMILARITY.
FT	DISULFID	240	282	BY SIMILARITY.
FT	DISULFID	271	295	BY SIMILARITY.
FT	DISULFID	303	434	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	346	362	BY SIMILARITY.
FT	DISULFID	354	423	BY SIMILARITY.
FT	DISULFID	448	523	BY SIMILARITY.
FT	DISULFID	480	496	BY SIMILARITY.
FT	DISULFID	513	541	BY SIMILARITY.
FT	CARBOHYD	153	153	POTENTIAL.
FT	CARBOHYD	187	487	POTENTIAL.
SO	SEQUENCE	566 AA:	63701 MW:	559A1E93B CRC32:

```

ID ACRO.RABIT STANDARD; PRT; 431 AA.
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE ACROSIN PRECURSOR (EC 3.4.21.10).
GN ACR
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE-TESTIS;
RX MEDLINE: 94368861.
RA RICHARDSON R.T., O'RAND M.G.;
  "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
  preproacrosin-related cDNA."
  Biochim. Biophys. Acta 1219:215-218(1994).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZO.
  IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
  IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
  ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
  CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation.
  CC the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05204; AAA61630.1;
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFAM: PF00089; trypsin; 1.
KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16
FT CHAIN 17 431
FT CHAIN 17 39
FT CHAIN 40 7
FT PROPEP ? 431
FT DISULFID 22 152
FT DISULFID 26 160
FT DISULFID 71 87
FT DISULFID 175 244
FT DISULFID 207 223
FT DISULFID 234 264
FT CARBOHYD 19 19
FT CARBOHYD 208 208
FT ACT_SITE 86 86
FT ACT_SITE 140 140
FT ACT_SITE 238 238
FT ACT_SITE 238 238
SO SEQUENCE 431 AA; 46422 MW; D350A173 CRC32;

Query Match 29.38; Score 555; DB 1; Length 431;
Best Local Similarity 39.18; Pred. No. 3,97e-113;
Matches 99; Conservative 54; Mismatches 81; Indels 19; Gaps 13;

```

```

QY 692 HYPAFLPLMRERPOXTASCTITGNG---DTGRATSRILQQAATPLLRKFE--ERIK 746
DB 215 GRITASNLCAGYPSG-K-IDTCGDSGGLMCRNENGEFFVVOGITSWGCARARPGI 272
QY 747 GRGTGMLCAGNLSHEKRPDSCGDSGGLPMC-ERRGESWVYGYVSMWCGCGVKNPCV 805
DB 273 YTAWPFPLDWIAS 285
QY 806 YTKVSAFVPIKWS 818

RESULT 11
ID ACRO.MOUSE STANDARD; PRT; 436 AA.
AC P23578;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACROSIN PRECURSOR (EC 3.4.21.10).
GN ACR
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91185335.
RA KASHIWABARA S., BABA T., TAKADA M., WATANABE K., YANO Y., ARAI Y.;
  "Primary structure of mouse proacrosin deduced from the cDNA sequence
  RT and its gene expression during spermatogenesis."
  J. Biochem. 108:785-791(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92041732.
RA WATANABE K., BABA T., KASHIWABARA S., OKAMOTO A., ARAI Y.;
  "Structure and organization of the mouse acrosin gene."
  J. Biochem. 109:828-833(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90255839.
RA KLEMM U., MAIER W.M., TSAOUSIDOU S., ADHAM I., WILLISON K.;
  "Mouse proacrosin: cDNA sequence, primary structure and
  RT postmeiotic expression in spermatogenesis."
  J. Biol. Chem. 265:16016-16021(1990).
RN [4]
RX MEDLINE: 92147126.
RA KREMLING H., KEIME S., WILHELM K., ADHAM I.M., HAMFISTER H.,
  ENGEL W.;
  "Mouse proacrosin gene: nucleotide sequence, diploid expression, and
  RT chromosomal localization."
  Genomics 11:828-834(1991).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZO.
  IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
  IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
  ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
  CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
  DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation.
  CC the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S66245; AAB20293.1;
DR EMBL: S64500; AAB20293.1; JOINED.

```

```

DR EMBL: S66243; AAB20293.1; JOINED.
DR EMBL: D00754; BAA00651.1; ALT_INIT.
DR EMBL: X52466; CAA36704.1; ALT_INIT.
DR EMBL: M85170; AAA40124.1; ALT_INIT.
DR EMBL: M96430; AAA37163.1; JOINED.
DR EMBL: M96427; AAA37163.1; JOINED.
DR EMBL: M96428; AAA37163.1; JOINED.
DR PIR: JX0172; JX0172.
DR HSSP: P00763; IDPO.
DR MGD: MGI:87884; ACR.
DR PROSITE: PS00134; TRYPSTIN_HIS; 1.
DR PROSITE: PS00135; TRYPSTIN_SER; 1.
DR PFM: PF00089; trypsin; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 19
FT CHAIN 20 436
FT CHAIN 20 42
FT CHAIN 43 7
FT PROPEP 436
FT DISULFID 25 155
FT DISULFID 74 90
FT DISULFID 178 247
FT DISULFID 210 226
FT DISULFID 237 267
FT CARBOHYD 22 22
FT CARBOHYD 211 211
FT ACT_SITE 89 89
FT ACT_SITE 143 143
FT ACT_SITE 241 241
FT ACT_SITE 241 241
FT CONFLICT 8 8
FT CONFLICT 10 14
FT CONFLICT 17 18
FT CONFLICT 23 23
FT CONFLICT 26 26
FT CONFLICT 33 33
FT CONFLICT 50 51
FT CONFLICT 122 122
FT CONFLICT 149 149
FT CONFLICT 173 173
FT CONFLICT 173 174
FT CONFLICT 177 183
FT CONFLICT 185 186
FT CONFLICT 188 189
FT CONFLICT 251 251
FT CONFLICT 251 258
FT CONFLICT 299 300
FT CONFLICT 320 320
FT CONFLICT 405 405
FT CONFLICT 407 436
FT CONFLICT 407 436
SQ SEQUENCE 436 AA: 48929 MW: 51ABE122 CRC32:
Query Match 28.98; Score 547; DB 1; Length 436;
Best Local Similarity 40.18; Pred. No. 5,44e-111;
Matches 101; Conservative 45; Mismatches 89; Indels 17; Gaps 13;

```

```

QY 693 VLPACPLWREPRQKTASNCYITGWC-DTGRA-Y-SRTIQAAIPILPRFCE--ERYNG 747
DB 219 RVTSTWVACAY-PEGR-IDICGDSGGLPCRDNDVSPVYVVGITSMVGACARAKRGVY 276
QY 748 RFTGMCLACANLHEHKKRVDSGCGSLMC--ERPESWVYGVTSWGYGCGVKNPSGVY 806
DB 277 TATWDYIDWTAS 288
QY 807 TKVSAPFVPIKS 818

RESULT 12
ID UROT_RAT STANDARD: PRT: 559 AA.
AC P19637;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TISSUE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.68) (TPA)
DE (T-PLASMINOGEN ACTIVATOR).
GN PLAT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA NY T., LEONARDSON G., HSUEH A.J.W.;
RT "Cloning and characterization of a cDNA for rat tissue-type
RT plasminogen activator."
RL DNA 7:671-677(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90130448.
RA FENG P., OHLSSON M., NY T.;
RT "The structure of the TPA-less rat tissue-type plasminogen activator
RT gene. Species-specific sequence divergences in the promoter predict
RT differences in regulation of gene expression."
RL J. Biol. Chem. 265:2022-2027(1990).
CC -1- FUNCTION: TPA CONVERTS PLASMINOGEN TO PLASMIN BY HYDROLYZING A
CC SINGLE R-V BOND IN PLASMINOGEN. ACTIVE IN TISSUE REMODELING AND
CC DESTRUCTION, PARTICULARLY IN FIBRINOLYSIS, AND IN CELL MIGRATION.
CC -1- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-1-VAL BOND IN
CC PLASMINOGEN TO FORM PLASMIN.
CC -1- MISCELLANEOUS: TPA ATTACHES TO THE KRINGLE STRUCTURE OF THE A
CC CHAIN OF FIBRIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M23697; AAA1812.1; -
DR EMBL: M31197; AAA42261.1; JOINED.
DR EMBL: M31185; AAA42261.1; JOINED.
DR EMBL: M31186; AAA42261.1; JOINED.
DR EMBL: M31187; AAA42261.1; JOINED.
DR EMBL: M31188; AAA42261.1; JOINED.
DR EMBL: M31189; AAA42261.1; JOINED.
DR EMBL: M31190; AAA42261.1; JOINED.
DR EMBL: M31191; AAA42261.1; JOINED.
DR EMBL: M31192; AAA42261.1; JOINED.
DR EMBL: M31193; AAA42261.1; JOINED.
DR EMBL: M31194; AAA42261.1; JOINED.
DR EMBL: M31195; AAA42261.1; JOINED.
DR EMBL: M31196; AAA42261.1; JOINED.

```


DR EMBL: A19618: CAA01482.1; -
 DR PIR: A31597; A31597.
 DR PIR: A35029; A35029.
 DR HSSP: P00750; 1RTF.
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01253; FIBRONECTIN_1; 1.
 DR PROSITE: PS00021; KRINGLE_1; 2.
 DR PROSITE: PS00070; KRINGLE_2; 2.
 DR PIR: P00008; EGF; 1.
 DR PIR: P00039; fn1; 1.
 DR PIR: P00051; kringle; 2.
 DR PIR: P00089; trypsin; 1.
 DR Plasminogen activation: Hydrolyase; Serine protease; Glycoprotein;
 Plasma; Kringle; EGF-like domain; Signal.
 SIGNAL 1
 PROPEP 18 29
 PROPEP 30 308
 CHAIN 309 559
 FT DOMAIN 36 78
 FT DOMAIN 79 117
 FT DOMAIN 124 205
 FT DOMAIN 213 294
 FT ACT_SITE 355 355
 FT ACT_SITE 404 404
 FT ACT_SITE 510 510
 FT DISULFID 38 68
 FT DISULFID 66 75
 FT DISULFID 83 94
 FT DISULFID 88 105
 FT DISULFID 107 116
 FT DISULFID 124 205
 FT DISULFID 145 187
 FT DISULFID 176 200
 FT DISULFID 213 294
 FT DISULFID 234 276
 FT DISULFID 265 289
 FT DISULFID 297 428
 FT DISULFID 340 356
 FT DISULFID 348 417
 FT DISULFID 442 516
 FT DISULFID 474 490
 FT DISULFID 506 534
 FT CARBOHYD 149 149
 FT CARBOHYD 481 481
 FT CONFLICT 380 380
 SEQUENCE 559 AA; 62903 MW; A106A89E CRC32;
 Query Match 28.9%; Score 546; DB 1; Length 559;
 Best Local Similarity 38.3%; Pred. No. 1.0e-110;
 Matches 93; Conservative 58; Mismatches 76; Indels 16; Gaps 13;

QY 819 VTK 821
 RESULT 13
 ID UROK_HUMAN STANDARD; PRT; 431 AA.
 AC P00749;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE UROKINASE-TYPE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.73) (UPA)
 DE (U-PLASMINOGEN ACTIVATOR).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85215647.
 RA RICCIO A., GRIMALDI G., VERDE P., SEBASTIO G., BOAST S., BLASI F.;
 RT "The human urokinase-plasminogen activator gene and its promoter."; Nucleic Acids Res. 13:2759-2771(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX HOLMES W.E., PENNICA D., BLABER M., REY M.W., GUENZLER W.A., STEFFENS G.J., HEYNEKER H.L.;
 RT "Cloning and expression of the gene for pro-urokinase in Escherichia coli."; Biotechnology 3:923-929(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86056954.
 RA NAGAI M., HIRAKATSU R., KANEDA T., HAYASUKE N., ARIMURA H., NISHIDA M., SUYAMA T.;
 RT "Molecular cloning of cDNA coding for human prepro-urokinase."; Gene 36:183-188(1985).
 RN [4]
 RP SEQUENCE OF 66-431 FROM N.A.
 RX MEDLINE: 84272706.
 RA VERDE P., STROPELLI M.P., GALEFFI P., DI NOCERA P., BLASI F.;
 RT "Identification and primary sequence of an unspliced human urokinase poly(A)+ RNA."; Proc. Natl. Acad. Sci. U.S.A. 81:4727-4731(1984).
 RN [5]
 RP SEQUENCE OF 21-177.
 RX MEDLINE: 83055084.
 RA GUENZLER W.A., STEFFENS G.J., OTTING F., KIM S.-M.A., FRANKUS E., FLOHE L.;
 RT "The primary structure of high molecular mass urokinase from human urine. The complete amino acid sequence of the A chain."; Hoppe-Seyler's Z. Physiol. Chem. 363:1155-1165(1982).
 RN [6]
 RP SEQUENCE OF 156-176 AND 179-224.
 RX MEDLINE: 83003608.
 RA SCHALLER J., NICK H., RICKLI E.E., GILLESSEN D., LERGIER W., STUDDER R.O.;
 RT "Human low-molecular-weight urinary urokinase. Partial characterization and preliminary sequence data of the two polypeptide chains."; Eur. J. Biochem. 125:251-257(1982).
 RN [7]
 RP SEQUENCE OF 158-410.
 RX MEDLINE: 83055099.
 RA STEFFENS G.J., GUENZLER W.A., OTTING F., FRANKUS E., FLOHE L.;
 RT "The complete amino acid sequence of low molecular mass urokinase from human urine."; Hoppe-Seyler's Z. Physiol. Chem. 363:1043-1058(1982).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE: 96000858.
 RA SPRAGON G., PHILLIPS C., NOMAK U.K., PONTING C.P., SAUNDERS D., DOBSON C.M., STUART D.I., JONES E.Y.;
 RT "The crystal structure of the catalytic domain of human

Search completed: Mon Mar 13 10:17:16 2000
Job time : 14 secs.

```

DR PIR: A29941; A29941.
DR HSSP: P00750; IRTF.
DR MGD: MGI:97610; PLAT.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 2.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PFAM: PF00008; EGF; 1.
DR PFAM: PF00039; fnl; 1.
DR PFAM: PF00051; kringle; 2.
DR PFAM: PF00089; trypsin; 1.
DR Plasma: Kringler; EGF-like domain; Signal.
KW Plasma: Kringler; EGF-like domain; Signal.
FT SIGNAL 1 17 PROBABLE.
FT PROPEP 18 29
FT CHAIN 30 308
FT CHAIN 309 559
FT DOMAIN 36 78 TISSUE PLASMINOGEN ACTIVATOR A CHAIN
FT DOMAIN 79 117 (HEAVY).
FT DOMAIN 124 205 TISSUE PLASMINOGEN ACTIVATOR B CHAIN
FT DOMAIN 213 294 (LIGHT, SERINE PROTEASE).
FT ACT_SITE 355 355 KRINGLE 1.
FT ACT_SITE 404 404 KRINGLE 2.
FT ACT_SITE 510 510 CHARGE RELAY SYSTEM.
FT ACT_SITE 510 510 CHARGE RELAY SYSTEM.
FT DISULFID 38 68 BY SIMILARITY.
FT DISULFID 66 75 BY SIMILARITY.
FT DISULFID 83 94 BY SIMILARITY.
FT DISULFID 88 94 BY SIMILARITY.
FT DISULFID 107 116 BY SIMILARITY.
FT DISULFID 124 205 BY SIMILARITY.
FT DISULFID 145 187 BY SIMILARITY.
FT DISULFID 176 200 BY SIMILARITY.
FT DISULFID 213 294 BY SIMILARITY.
FT DISULFID 234 276 BY SIMILARITY.
FT DISULFID 265 289 BY SIMILARITY.
FT DISULFID 297 428 INTERCHAIN (BY SIMILARITY).
FT DISULFID 340 356 BY SIMILARITY.
FT DISULFID 348 417 BY SIMILARITY.
FT DISULFID 442 516 BY SIMILARITY.
FT DISULFID 474 490 BY SIMILARITY.
FT DISULFID 506 534 BY SIMILARITY.
FT CARBOHYD 149 149 POTENTIAL.
FT CARBOHYD 481 481 POTENTIAL.
SO SEQUENCE 559 AA; 63110 MW; CEEAF10F CRC32;

27.9%; Score 527; DB 1; Length 559;
Local Similarity 38.3%; Pred. No. 1.16e-105;
Matches 93; Conservative 56; Mismatches 78; Indels 16; Gaps 14;

```

 NWSE (7M)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

h-pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 13 10:17:33 2000; Maspar time 29.51 Seconds

Tabular output not generated.

Title: >US-09-147-947-6
 Description: (578-822) from US09147947A.pep (1 of 6)
 Perfect Score: 1892
 Sequence: 1 IIGGRNSLRGMPQVSLRL.....PGVTKSAFVPIKSVTKL 245

Scoring table: PAM 150
 Gap 11

Searched: 225878 segs, 69334122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: ptrembl12
 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
 13:sp.vetebtrate 14:sp.virus

Statistics: Mean 43.807; Variance 67.987; scale 0.644
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	ID	Description	Pred. No.
1	625	33.0	571	4 Q9Y495	PLATELET FACTOR XI.	1.12e-127
2	620	32.8	868	5 Q9Y1V3	TUNICATE RETINOIC ACID	2.25e-126
3	588	31.1	581	5 Q9XZM7	CORTICAL GRANULE SERIN	4.81e-118
4	574	30.3	415	6 Q29015	PREPROGROSIN.	2.07e-114
5	558	29.5	377	6 P79343	ACROSIN.	2.88e-110
6	554	29.3	1042	4 Q9Y505	CORN.	3.12e-109
7	545	28.8	421	11 Q60491	PREPROGROSIN PRECURSO	6.61e-107
8	541	28.7	411	4 Q15844	UROKINASE-TYPE PLASMIN	7.12e-106
9	541	28.6	560	4 Q14520	HGF ACTIVATOR LIKE PRO	7.12e-106
10	542	28.6	1113	11 Q92319	LOW DENSITY LIPOPROTEI	3.93e-106
11	535	28.3	806	6 Q18783	PLASMINOGEN.	2.51e-104
12	530	28.0	431	4 Q16618	UROKINASE PRECURSOR (E	4.88e-103
13	530	28.0	643	6 Q97506	KALIKREIN.	4.88e-103
14	518	27.4	297	11 Q88781	SERINE PROTEASE PRECUR	5.98e-100
15	512	27.1	1524	13 Q91674	POLYPROTEIN (OVOCYTHAS	2.08e-98
16	507	26.8	418	4 Q60235	AIRWAY TRYPSIN-LIKE PR	4.00e-97
17	503	26.6	283	4 Q95518	DJ1170K4.2 (NOVEL TRYP	4.25e-96
18	489	25.8	260	13 Q9W799	ELASTASE 4 PRECURSOR (1.63e-92
19	488	25.8	261	13 Q9W704	CHYMOTRYPSINOGEN 1.	2.94e-92
20	489	25.8	334	6 Q46507	PLASMINOGEN (FRAGMENT)	1.63e-92

PLASMINOGEN PRECURSOR. 1.72e-91
 PLASMINOGEN ACTIVATOR 1.05e-89
 FACTOR I C3B/C4B INACT 1.11e-88
 APOLIPOPROTEIN A (FRAG 6.43e-88
 SIMILAR NO PERTIDASE F 1.16e-87
 SP0011A (FRAGMENT). 6.72e-87
 COMPLEMENT COMPONENT F 3.90e-86
 FXII. 3.90e-86
 SERINE PROTEINASE (FRA 1.26e-85
 FACTOR I PROTEIN PRECU 1.26e-85
 SERINE PROTEINASE. 2.52e-82
 TRYPTEASE-111 (FRAGMENT 4.66e-81
 SERINE PROTEASE (BSP). 1.50e-80
 ELASTASE PRECURSOR (EC 8.59e-80
 TRYPTASE-1. 8.59e-80
 MYLENEPEPALON SPECIFI 8.83e-79
 FACTOR XII. 2.83e-78
 OVIDUCTIN. 5.06e-78
 TRYPISINOGEN 1. 5.17e-77
 FACTOR X. 9.43e-76
 SERINE PROTEASE (FRAG 9.43e-75
 TRYPISINOGEN 2 (FRAGEN 1.68e-75
 ELASTASE 3 PRECURSOR. 1.68e-75
 TRYPISIN-LIKE SERINE PR 1.68e-75
 LIMULUS FACTOR D. 9.59e-75

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	571 AA.
ID	Q9Y495			
AC	Q9Y495;			
DT	01-NOV-1999 (TREMREL. 12, Created)			
DT	01-NOV-1999 (TREMREL. 12, Last sequence update)			
DT	01-NOV-1999 (TREMREL. 12, Last annotation update)			
DE	PLATELET FACTOR XI.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98256306.			
RA	HU T.C.; SHORE S.K.; SESHMA T.; BAGASRA O.; WALSH P.N.;			
RT	"Molecular cloning of platelet factor XI, an alternative splicing			
RT	product of the plasma factor XI gene."			
RL	U. Biol. Chem. 273:13787-13793(1998).			
DR	EMBL; AF045649; AAC24506.1; -			
DR	PROSITE; PS00495; APPE; 3.			
DR	SEQUENCE 571 AA; 63840 MW; C25DBIA9 CRC32;			

Query Match 33.0%; Score 625; DB 4; Length 571;
 Best Local Similarity 41.88; Pred. No. 1.12e-127;
 Matches 102; Conservative 53; Mismatches 71; Indels 18; Gaps 16;

Db	334	IVGGTASVREMPVYLH-TTSEPTQ-RHLCGSGIINOMITRAHCF--YGVESPILR 389
Qy	578	IIGGRNSLRGMPQVSLRLKSSHGDRLLCGATLLSSCVLTAHCFKRYG-NSTRSYA 636
Db	390	VYSGILNOSKED-TSEFVQELIHDQYKMAESGYDIALLKLETTVN-YT--DSOR-P 444
Qy	637	VVRGVD-YHTLVPEFEFEIEIVQOIVIREYRPDRSDIDIALVRLOGDEECARFSSVLP 695
Db	445	ICLPSKGR-NVIYTDQVWGMGRKRLDQIOMTLQAKAPLYVNECCQRYRKHITKH 503
Qy	696	ACPLWERPQKTSNLYITGWC-DTGRA-YSTRLOQAAPLPKRCERRYG-RTGR 752
Db	504	MICAGY-REGGK-DACKGDSGGLPSC-KHNEVHLVIGITSGSCAORERPGYTNVEY 560
Qy	753	MLCAGNHNHKKRVDSCGDSGGLMCEPESNVVYGVTSWGIGCGKSDPGYTKSAF 812
Db	561	VDWT 564
Qy	813	VPWT 816

Protease
Att No. 9 is acid stable

RESULT	2	PRELIMINARY:	PRT:	868 AA.
ID	09Y1V3			
AC	09Y1V3:			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	TUNICATE RETINOIC ACID-INDUCIBLE MODULAR PROTEASE PRECURSOR.			
GN	TRAMP.			
OS	Polyandrocarpa misakiensis.			
OC	Eukaryota; Metazoa; Chordata; Urochordata; Ascidacea;			
OC	Stolidobranchia; Styelidae; Polyandrocarpa.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-WHITE SPOT;			
RA	OHASHI M., KAWAMURA K., FUJII N., YUBISUI T., FUJIMARA S.;			
RE	"A retinoic acid-inducible modular protease in budding ascidians.";			
RE	Dev. Biol. 0:0-0(1999).			
PROSITE:	PS01209; LDLRA_1: 3.			
SIGNAL	Protease; glycoprotein.			
FT	SIGNAL	1	28	POTENTIAL.
SQ	SEQUENCE	868 AA;	97660 MM;	682828AE6 CRC32;
Query Match	32.8%;	Score 620;	DB 5;	Length 868;
Best Local Similarity	39.5%;	Pred. No. 2,25e-16;		
Matches	100;	Conservative	56;	Mismatches 77; Indels 20; Gaps 13.
Db	624	IVGGSGTEPHENPQAGIWLPMY---N--CGGSIIHCWLTAAHCFVR-EYPIRDYTI	677	
Oy	578	IIGGKNSLRGGPWPQVSLRKSSHDDGLLCATILSSCWLTAAHCKRGNSIRSYAV	637	
Db	678	RLGDHITGVDETEQLFKIAELIKH-DYVYTKENDIALRLIENDARECATITPEVOYVC	736	
Oy	638	RVGDHITVLPDEFEEIEIGVOQIVIHREYRPRSDYDIALVRLQGEQCAFSSHLVLPAC	697	
Db	737	LP--KSSSOFPAKTICEYTGWCKSDATAVRAVYVPLQAEPLPLANKKCLDSEYT-QLG	793	
Oy	658	LPWMEREPD-KRASCXYITGWG-D--TG-RVRSRTLDQAAIPLPKRC--EERIKGRFT	750	
Db	794	PTMFCAGYLTGK--DSCGDSGGPILSCRDSDDRYYVWGVISWNGCGCAKPKAGVYAKV	851	
Oy	751	GMILCAGMLHEHKRVDSQGDGGPILMC-ERRGEWVYVGYTSMYGCGVADSPGVYTKV	809	
Db	852	AVFDIMIQMTGL	864	
Oy	810	SAFVPWIKSVTKL	822	
FT	3			
FT	09XZM7	PRELIMINARY:	PRT:	581 AA.
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	CORTICAL GRANULE SERINE PROTEASE 1 PRECURSOR.			
OS	Strongylocentrotus purpuratus (Purple sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;			
OC	Euechinozoa; Echinoidea; Echinozoa; Strongylocentrotidae;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	HALEY S.A., WESSLE G.M.;			
RT	"The cortical granule serine protease CGSP1 of the sea urchin,			
RL	Strongylocentrotus purpuratus, is autocatalytic and has an LDL			
RL	receptor-like domain.";			
RE	Dev. Biol. 0:0-0(1999).			
RE	EMBL: AF149789; AAD37426.1: 3.			
DR	PROSITE: PS01209; LDLRA_1: 3.			
FT	SIGNAL: Protease; glycoprotein.			
FT	SIGNAL	1	29	POTENTIAL.
FT	CHAIN	30	581	CORTICAL GRANULE SERINE PROTEASE 1.

SQ	SEQUENCE	581 AA;	64438 MW;	3CA00786 CRC32:
	Query Match	31.1%;	Score 588;	DB 5; Length 581;
	Best Local Similarity	40.0%;	Pred. No. 4.81e-118;	
	Matches	98;	Conservative	55; Mismatches 78; Indels 14; Gaps 12.
Dd	334 IYGGGPATAGMPWMOAOEFLRT-RSOMLVCGGTLIDPOVLTAAHCFMPMATSRNOY	392		
Oy	578 IIIGGNLSRLGGNPNOWSVLRKSSHQDGLLGCATLLSCSWVLTAHCRRKYNSTRSYAV	637		
Dd	393 HLGKHSVDVEPAGSOHLRVREIFVKFKFGHGVCGDIALILIDEVPQ-E-TGOINW	449		
Oy	638 RUGDYHT-LVPEEFEEIEYGQOIVIHREYRDRS-DVIDALVRLOGEEOCARSSHLP	695		
Dd	450 ACLD--EGMPLNDRECTISGWVTGMGNGPDYLHEARMLPRICNTKKSTNGRIEK	507		
Oy	696 ACPLMRERPORTASNCYTITMGDGTRA-YSR-TLOQAIPILLKEKRCE-ER-YKGFTG	751		
Dd	508 TWLCAGH-EGG-IDACGGSGSPCLSGPDPHNVGVTSWGHCIAANKPGYTVSS	565		
Oy	752 RMLCAGNHCHKRDVCCGGSDGSLMCERPEESMVYIGVISWIGCVGSFYTRVSA	811		
Dd	566 YLDWI 570			
Oy	812 FYPWI 816			
	RESULT	4		
ID	Q29015 PRELIMINARY; PRT; 415 AA.			
AC	Q29015;			
DT	01-NOV-1996 (TRENBLREL. 01. Created)			
DT	01-NOV-1996 (TRENBLREL. 01. Last sequence update)			
DE	01-NOV-1999 (TRENBLREL. 12. Last annotation update)			
OC	PREPROACROSIN.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
RN	Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
RP	[1]			
RC	SEQUENCE FROM N.A..			
RX	TISSUE=LIVER;			
RA	MEDLINE: 96348713.			
RA	AADAM T.M., KREMELING W., NIETTER S., ZIMMERMAN S., HUMMEL M.,			
RT	SCHROETER U., ENGEL W."			
RT	"The structures of the bovine and porcine proacrosin genes and their			
RL	conservation among mammals.";			
DR	Biol. Chem. Hoppe-Seyler 377:261-265(1996).			
DR	EMBL: X58549; CAA1440.1; -			
DR	PFAM: PF00089; trypsin.1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
FT	CHAIN 17 415 ACROSIN.			
SO	SEQUENCE 415 AA; 45361 MW; 80A12323 CRC32;			
	Query Match : 30.3%; Score 574; DB 6; Length 415;			
	Best Local Similarity: 39.1%; Pred. No. 2.07e-114;			
	Matches 95; Conservative 56; Mismatches 79; Indels 19; Gaps 14.			
Dd	40 VVGGAAPGAWPVWVSLOIFMYHNRRHYTCGAILNSHWVLTAAFCFNKKKYTDMLR	99		
Oy	578 IIIGKNLSRLGGNPWYSRLKSSHQDGL-LCGATLLSSCWVLTAHCRRKYNSTRSYAV	637		
Dd	100 IFGANVVWSNKPYPPLOGRF-VETIIHKYVSGLEINDIALIKI-TPVPACGF-	155		
Oy	634 SYAVR-VG-DYHTLVPEEFEIEIGVOQLVIHREYRPRSDDIALVRLROGEEOCARFS	691		
Dd	156 -IGPCPLQFOFAAGPPRARQCWTGTGWLKRGKGRPTSTLOEAVALIDELCNSTRWN	214		
Oy	692 HVLPACPCLMRBRPKXTASNCTIYWGG--DTGRAYSTILOOAIPILPRFCEE-R-YK	746		
Dd	215 GRIRATNYCACGYPR-K-IDTCOGSGGPLMCRORAEITFWVAITSMTGVCACAKRAPY	272		
Oy	747 GRTFRMLCAGNLHHKHKVYDSGCGDSGGLPMC-EPGESWVYIVGTSMGIGCYKXDSPGV	805		
Dd	273 YTSTWPYLNWIAS 285			

QY 806 YTKVSAFVPMIKS 818

RESULT 5 PRELIMINARY: PRT: 377 AA.

AC P79343;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE ACROSIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.

SEQUENCE FROM N.A.

TISSUE-LIVER;

KEIME S.;
Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

TISSUE-LIVER;

ADHAM I.M., KREMLING H., NIETTER S., ZIMMERMANN S., HUMMEL M.,
SCHROETER U., ENGEL W.;
"The structures of the bovine and porcine proacrosin genes and their
conservation among mammals";
Biol. Chem. Hoppe-Seyler 377:261-265(1996).

EMBL: X68212; CAA48294.1; -.

HSSP: P00763; IDPO.

PFAM: PF00089; trypsin; 1.

PRINTS: PR00722; CHYMOTRYPSIN.

SEQUENCE 377 AA; 41722 MW; E646321D CRC32;

Query Match 29.58; Score 558; DB 6; Length 377;
Best Local Similarity 37.9%; Pred. No. 2.88e-110;

Matches 96; Conservative 56; Mismatches 82; Indels 19; Gaps 15;

Db 3 IIGGGAAGSWMWVSLDIFTHNNRRYVCGLLNAHMLTAHCFIRKKYVDWL 62

QY 578 IIGGKNSLRGMPWVSLKSSHGGRLLCGATLLSSCWLTAAHCF--RYGNST-R- 633

Db 63 IFGAEVEMGSKRPVPLQERY-VEKIIHEKYSASSEANDIALIKITPPVY-CGPF-- 118

QY 634 SYAVR-VG-DYHTLVPEEFEEELGVQIYIHREYRDRSDYDALVRLOGPEEQCARFS 691

Db 119 -IGPGCLPFRAGPRPVYPCVWAGVFLERARRTSPVLQEAHVDLIDLDLNCSTRWYN 177

QY 692 HVLPLCLPLMRERPOKTASNCYITGNG--DTGRAYSRTLQQAIPLLPRFCEE-R-YK 746

Db 178 GRIKRSINVCAGY-PEKG-IDTCGGSGGGLMKDSVNSVYVVGISWVGCSRRARPGV 235

QY 747 GRTGRLCLGNNHEHKKRVDSGCGSGGLMCC-ERPEGSWVYGVTSWVGCGVCKDSPG 805

Db 236 YTSTWSYLNWIAS 248

QY 806 YTKVSAFVPMIKS 818

RESULT 6 PRELIMINARY: PRT: 1042 AA.

AC Q9Y505;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE CORIN.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;

RX MEDLINE: 99262646.
RA YAN W., SHENG N., SENO M., MORSE J., WU Q.;
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
from human heart";
RT J. Biol. Chem. 274:14926-14935(1999).

DR EMBL: AF133845; AAD31850.1; -.

DR PROSITE: PS01209; LDLRA_1; 6.

KW Glycoprotein.

SEQUENCE 1042 AA; 116565 MW; FE3227D1 CRC32;

Query Match 29.38; Score 554; DB 4; Length 1042;
Best Local Similarity 40.5%; Pred. No. 3.12e-109;

Matches 98; Conservative 43; Mismatches 87; Indels 14; Gaps 10;

Db 802 IIGGRTSRGPRWPCSLD--SEPS-GHI-CGCVLAKKWTVAHCFEGRENA-VMKY 856

QY 578 IIGGKNSLRGMPWVSLKSSHGGRLLCGATLLSSCWLTAAHCFIRKYSTSYAV 637

Db 857 VGINNLDHPVFMQTRFKTIIHPRYSRAVVDISTVEL--SEDISE--TGYPVPC 912

QY 638 RVGDYHTLVPEEFEEELGVQIYIHREYRDRSDYDALVRLOGPEEQCARFSHVLPA 697

Db 913 LPR-PEQMLEPDTCTCTGSHGKMKPFKLOGEYRITSLHCOSYFPMKITTMTICA 971

QY 698 LPLMRERPOKTASNCYITGNGDTRAYSRTLQQAIPLLPRFCEEYRGR-FTGRMLCA 756

Db 972 G--YESGTYDSCGDSGGGLVCEKPGRWTLFLGTFSGSVCSKVLGPGVSNVSYFEM 1029

QY 757 GNLHEKRRVDSGCGSGGLMCCERPEGSWVYGVTSWVGCGVKD-SPGVYTKVSAFVPM 815

Db 1030 IK 1031

QY 816 IK 817

RESULT 7 PRELIMINARY: PRT: 421 AA.

AC Q60491;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)

DE PREPROACROSIN PRECURSOR (FRAGMENT).

OS Cavia porcellus (Guinea pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RA GERTON G.L., HOFF H.B., BABA T.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z12153; CAA78137.1; -.

DR HSSP: P00763; IDPO.

DR PFAM: PF00089; trypsin; 1.

DR Signal.

FT NON_TER

FT SIGNAL

FT CHAIN

FT SIGNAL

FT CHAIN

FT SIGNAL

FT CHAIN

FT SIGNAL

FT CHAIN

FT SIGNAL

FT CHAIN

FT SIGNAL

FT CHAIN

FT SIGNAL

Oy	652	HVLPLCEPLMEEREPKPTASNCYITGWSGT-GRAY-STTLQQAAILPLPKFCE--RRYK	746
Db	216	GRVSTNVACQ-PRGK-IDTCOGSGSGELMCRDANSPFYVYITSGVCAAKRPGI	273
Oy	747	GRFGRMICACGLNHHKRNVDSCQGDSGPLMC-ERPESWVYIVGTSMGICCGYKDEPGV	805
Db	274	YTATMDYLDWIAS	286
Oy	806	YTKVSAFVPMIKS	818

RESULT	8			
ID	Q15844	PRELIMINARY:	PRT:	411 AA.
AC	Q15844:			
DT	01-NOV-1996	(TREMBLrel. 01,		Created)
DI	01-NOV-1996	(TREMBLrel. 01,		Last sequence update)
DT	01-NOV-1999	(TREMBLrel. 12,		Last annotation update)
TI	UROKINASE-TYPE PLASMINOGEN ACTIVATOR (FRAGMENT).			
UPA	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
NC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 86056954.			
RA	NAGAI M., HIRAMATSU R., KANEDA T., HAYASUKE N., ARIURA H.,			
RT	NISHIDA M., SUYAMA T.: "Molecular cloning of cDNA coding for human preprourokinase."			
RL	Gene 36:183-188(1985).			

RP SEQUENCE FROM N.A.
RX MEDLINE: 96186279.
RA YOSHIMOTO M., USHITAMA Y., SAKAI M., TAMAKI S., HARA H., TAKAHASHI K.,
RA SANASAKI Y., HANADA K.,
RT "Characterization of single chain urokinase-type plasminogen activator
RT with a novel amino-acid substitution in the kringle structure."
RL Biochim. Biophys. Acta 1293:83-89(1996).
DR HESB: D11443; BAA01919.1; -
DR HESP: P00749; IURK.
DR PFAM: PF00051; kringle1.1.
DR Pfam: PF00089; trypsin; 1.
LN Serine protease.
KW
FT NON_TER 1
FT CONFLICT 121 121 L -> P (IN REF. 1).
SQ SEQUENCE 411 AA; 46384 MW; 254CBB0E CRC32;

Query March 28.7% Score 543; DB 4; Length 411;
Best Local Similarity 36.9%; Pred. No. 2,17e-106;
Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;

Q7	578	11GKNSJLGGCWPQVOSJRLSSHGDRLLCGATLLSCWVLTAAHCFKRYGNSTREYAV	6377
Db	218	Y1GSRJLNSNQGEMKFEVETLLHKRYSDAPTLAHNHDILNKRKREGCACPSRTIOT	2777
QY	638	RVGHYHLLVPEEFEEJGVQOIVYHREPRDR-SDY-DIALRQGEDEECAPFFSHLP	6955
Db	278	ICLPSMYND-PQ-EGTSCEITGFGKENSTDYLYPEOLKMTVYLIHSRECOQPHYYGSEV	3355
QY	696	ACLR-LMRERPOKTAASNCYIITGWC-DIGRAY-SRTLQQAIPLLPRFCEE-RYKGR-F	7495
Db	336	TTKMLCAAD-PQWK-TSCGCGDSGLPYVCSLQGM-1LTGIVSMGRCCALKDKPQYTRK	3925
QY	750	TGRRLCAGNLHEHRVYDSCGDSGGLPMDCEPGESEWVYGVTSWIGVGGVXKDSPVYTRK	8095
Db	393	SHLPWIRSHTK	4045
QY	810	SAFVPWIKSVTK	8215
RESULT	9		
ID	Q14520	PRELIMINARY;	PRT; 560 AA.

014520: 000663;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HGF ACTIVATOR LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA KITAHARA N.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96425001.
RA CHOI-MINGA N.H., TOBE T., SUMIYA J., NAKANO Y., SANO Y., MADA T.,
RA TOMITA M.;
RT "Purification and characterization of a novel hyaluronan-binding
RT protein (PHP) from human plasma: it has three EGF, a kringle and a
RT serine protease domain, similar to heptocyte growth factor
RT activator".
JL Biochem. 119:1157-1165(1996).
DR EMBL: DA9742; AAA08576.1;
DR EMBL: S83182; AAB86909.1; -;
DR HSSP: P00763; IDPO.
DR PFAM: PF00008; EGF_3.
DR PFAM: PF00051; kringle; 1.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PR00018; KRINGLE.
DR PRINTS: PR00722; CHYMOTRYPSIN.
SQ SEQUENCE 560 AA; 62671 MW; 4AC81907 CRC32;

Query Match	28.6%	Score 541	DB 4	Length 560
Best Local Similarity	41.8%	Pred. No. 7,12e-106		
Matches	99	Conservative	45	Mismatches 73
			Indels 20	Gaps 14
Db	332	QSLEPLISMPOGHF-CGGAIIHPCWLTAAHC-T-DIKTRHLKVLVGD-QDLKKEEFH	386	
Oy	592	QVSLRKSSHDGDRLLCATLLSCWVLTAAHCKRRKGNSTRSAVAVRHDTLVLPREE	651	
Db	387	EOSFVEKIEFKSMYNENDEIIPHNDIALIKLPRDGCALCESKTYVKVYCLP--DSSEFS	443	
Oy	652	EE-IGVQDIVIHREI--RPDRSDYDIALVRLQGEEDCARESSHVLPACLPLMERPOKT	708	
Db	444	GSECHISGMVLTETGK-GSRQLDAKVKLIINILCNSROLTDHMKIDDSMLCAGNLIQRPGQ	502	
Oy	709	ASNCYITGMG--DGRAYSRTLQDAIPLRLPKRCSEER-YKGFRTGMLCAGNLEHKKR	764	
Db	503	-DTQGDGSGGLTLEKDG-TYYVVGIVSMGLECG-K-RPGVYTYOVTFLFKIRATIK	555	
Oy	765	VDSCQGDGSGGLMERPEASVWVGVGVISWGGCVKSPGYITTVKSAFVPIKSVTK	821	

RESULT	10		
ID	Q9Z319	PRELIMINARY;	PRT; 1113 AA.
AC	Q9Z319;		
DT	01-MAY-1999	(TEMBREL. 10, Created)	
DT	01-MAY-1999	(TEMBREL. 10, Last sequence update)	
DT	01-NOV-1999	(TEMBREL. 12, Last annotation update)	
DE	LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98429596.		
RA	TOMITA Y., KIM D.-H., MAGOORI K., FUJINO T., YAMAMOTO T.T.;		
RT	A novel low-density lipoprotein receptor-related protein with type III		
RT	membrane protein-like structure is abundant in heart.,"		
RL	J. Biochem. 124:784-789(1998).		
DR	EMBL; AB013874; BAA34371.1; -.		
DR	HSSP; P00763; IDPO.		
DR	PROSITE; PS01209; LDLRA_1; 6.		

(WL)

5

e-45

00e-45

QY 818 SVTKL 822

RESULT 2

ID W83361 standard; Protein: 875 AA.

AC W83361:

DT 17-FEB-1999 (first entry)

DE Human neurotrophin.

KW Human; neurotrophin; tumour inhibition; neurological disease;

KW lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

KW emphysema; bronchitis.

OS Homo sapiens.

PN W09849322-A1.

PD 05-NOV-1998.

PF 24-APR-1998; IB0625.

PT 26-APR-1997; CH-000966.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

PT New human and murine neurotrophin - used, e.g. for inhibiting

PT tumours, treatment of neurological or lung disease, including by

PT gene therapy and in drug development

PS Claim 1; Page 20-24; 50pp; English.

CC The present sequence represents human neurotrophin. Neurotrophin proteins

CC and polynucleotides can be used: (i) to inhibit tumours, including

CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage

CC caused by stroke or brain injury (having a protective effect on the

CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

CC increase survival of damaged neurons (e.g. in cases of hypoxia,

CC ischaemia, nerve transection) and to stimulate regeneration and/or

CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

CC degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

CC causes of cell death) in the nervous system; (vii) to regenerate brain

CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

CC performance, including learning and memory; (x) to treat or prevent a

CC wide range of psychiatric disorders; and (xi) to treat brain or lung

CC injury associated with protease expression (specifically emphysema or

CC bronchitis).

SQ Sequence 875 AA.

Query Match 100.0%; Score 1892; DB 1; Length 875;

Best Local Similarity 100.0%; Pred. No. 9,99e-193;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 631 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSRYAV 690

QY 578 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSRYAV 637

DB 691 RVGDHTLVPEEFEEIGVQOVIHREYRPRDSYDIALVRLQGPEDCARFSSHYLPAC 750

QY 638 RVGDHTLVPEEFEEIGVQOVIHREYRPRDSYDIALVRLQGPEDCARFSSHYLPAC 697

DB 751 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAVPLPKRCEERYKRGFTGRLACAG 810

QY 698 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAVPLPKRCEERYKRGFTGRLACAG 757

DB 811 NLHEKRVDSGCGDGGPLMCEPESWVYVGTSMGCGGVADPGVYTKVSAFPMWK 870

QY 758 NLHEKRVDSGCGDGGPLMCEPESWVYVGTSMGCGGVADPGVYTKVSAFPMWK 817

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

QY 818 SVTKL 822

DB 871 SVTKL 875

DE Mouse neurotrophin.

KW Mouse; neurotrophin; tumour inhibition; neurological disease;

KW lung disease; gene therapy; drug development; stroke; brain injury;

KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;

KW epilepsy; hypoxia; ischaemia; nerve transection; neurogenesis;

KW emphysema; bronchitis.

OS Mus musculus.

PN W09849322-A1.

PD 05-NOV-1998.

PF 24-APR-1998; IB0625.

PT 26-APR-1997; CH-000966.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

DR N-PSDB: V72590.

PT New human and murine neurotrophin - used, e.g. for inhibiting

PT tumours, treatment of neurological or lung disease, including by

PT gene therapy and in drug development

PS Claim 1; Page 29-32; 50pp; English.

CC The present sequence represents mouse neurotrophin. Neurotrophin proteins

CC and polynucleotides can be used: (i) to inhibit tumours, including

CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage

CC caused by stroke or brain injury (having a protective effect on the

CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

CC increase survival of damaged neurons (e.g. in cases of hypoxia,

CC ischaemia, nerve transection) and to stimulate regeneration and/or

CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

CC degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

CC causes of cell death) in the nervous system; (vii) to regenerate brain

CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

CC performance, including learning and memory; (x) to treat or prevent a

CC wide range of psychiatric disorders; and (xi) to treat brain or lung

CC injury associated with protease expression (specifically emphysema or

CC bronchitis).

SQ Sequence 761 AA.

Query Match 92.6%; Score 1752; DB 1; Length 761;

Best Local Similarity 89.8%; Pred. No. 3.43e-177;

Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;

DB 517 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSRYAV 576

QY 578 IIGKNSLGGMPWQVSLKSSHGDRLLCGATLLSSCWVLTAAHCFKRGNSRYAV 637

DB 577 RVGDHTLVPEEFEEIGVQOVIHREYRPRDSYDIALVRLQGPEDCARFSSHYLPAC 636

QY 638 RVGDHTLVPEEFEEIGVQOVIHREYRPRDSYDIALVRLQGPEDCARFSSHYLPAC 697

DB 698 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAVPLPKRCEERYKRGFTGRLACAG 757

QY 637 LPLMRERPKTASNCYITGMDTGRAYSRITLQAAVPLPKRCEERYKRGFTGRLACAG 696

DB 697 NLQEDNRVDSGCGDGGPLMCEPESWVYVGTSMGCGGVADPGVYTKVSAFPMWK 756

QY 758 NLHEKRVDSGCGDGGPLMCEPESWVYVGTSMGCGGVADPGVYTKVSAFPMWK 817

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

DB 757 SVTKL 761

QY 818 SVTKL 822

RESULT 4

ID W99088 standard; Protein: 761 AA.

AC W99088:

DT 13-MAY-1999 (first entry)

DE Mouse serine protease BSSP-3.

KW Serine protease; BSSP-3; brain tissue.

OS Mus sp.

PN W09905290-A1.

PD 04-FEB-1999.

PF 24-JUL-1996; J03324.

PT 24-JUL-1997; JP-213969.

PS (SOND/) SONDEREGGER P.

PT Sanderegger P.

PI: 99-009438/01.

DR N-PSDB: V72590.

PT New human and murine neurotrophin - used, e.g. for inhibiting

PT tumours, treatment of neurological or lung disease, including by

PT gene therapy and in drug development

PS Claim 1; Page 29-32; 50pp; English.

CC The present sequence represents mouse neurotrophin. Neurotrophin proteins

CC and polynucleotides can be used: (i) to inhibit tumours, including

CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage

CC caused by stroke or brain injury (having a protective effect on the

CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to

CC increase survival of damaged neurons (e.g. in cases of hypoxia,

CC ischaemia, nerve transection) and to stimulate regeneration and/or

CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.

CC degeneration or neurodegeneration); (vi) to prevent apoptosis (or other

CC causes of cell death) in the nervous system; (vii) to regenerate brain

CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain

CC performance, including learning and memory; (x) to treat or prevent a

CC wide range of psychiatric disorders; and (xi) to treat brain or lung

CC injury associated with protease expression (specifically emphysema or

CC bronchitis).

SQ Sequence 761 AA.

OS Synthetic.
 FH Key Location/Qualifiers
 FT region 1..10
 FT /note= "residues 1-10 of native hPUK"
 FT region 11..389
 FT /note= "residues 33-411 of native hPUK"
 FT misc_difference 155
 FT /note= "corresponds to TAC codon"
 PN EP-398361-A.
 PD 22-NOV-1990.
 PE 18-MAY-1990: 109472.
 PR 03-JUN-1986: JP-156936.
 PR 18-FEB-1987: JP-036495.
 PR 18-MAY-1989: JP-126434.
 PR 18-MAY-1989: JP-126433.
 PR 22-FEB-1990: JP-042020.
 (GREG, GREEN CROSS CORP.
 Airmura H, Amatsuji Y, Hirose M, Kasai S, Kawabe H;
 Morita M, Tanabe T;
 DR WPI; 90-350146/47.
 DR N-PSDB; T61673.
 PT Human pro-urokinase variants - deficient in loop regions of
 PT epidermal growth factor, showing long blood half-life, as
 PT fibrinolytic agent
 PS Claim 6; Page -; 22pp; English.
 CC New variants of human prourokinase (hPUK) comprise a hPUK deficient
 CC in (i) at least part of the first loop region of the epidermal growth
 CC factor (EGF) domain; (ii) at least part of the first loop and at
 CC least part of the second loop; or (iii) at least part of the third
 CC loop. The hPUK variants show an increased blood half-life comparable
 CC to that of the whole EGF domain-deficient hPUK variant and urokinase
 CC while retaining the same properties as those of hPUK. They have
 CC potent thrombolytic activity and very little tendency to cause
 CC spontaneous bleeding. The present sequence represents a specific
 CC variant of hPUK which lacks loops 1 and 2 of the EGF domain; the
 CC sequence does not appear in the specification and has been created
 CC using the wild-type hPUK sequence and the junction sequence after
 CC deletion, both of which are given (in Fig 1 and on page 8,
 CC respectively).
 SQ Sequence 389 AA;

Query Match 29.0%; Score 548; DB 1; Length 389;
 Best Local Similarity 36.9%; Pred. No. 3.35e-45;
 Matches 93; Conservative 60; Mismatches 85; Indels 14; Gaps 13;
 Db 137 IIGGEFTIENQPFATIRRRHGGSVIYVGGSLISPCWISATHCIDPKKE-DYIV 195
 578 IIGKNSLRGWPQVQSLRKS SHGDGRLCGATLLSCWVLTAAHCFKRYGNSTRSYAV 637
 196 YLGRSLNSNTGEMKEFEVENTLHKDYADTLAHNDIALKIRSEKRCQPSRTIQT 255
 QY 638 RVGYHTLVPEFEELICVQIVIHREYRPR-SDY-DIALVRLQGPEDQCARFSSHYLP 695
 Db 256 ICLPSMTND-PQ-FGTSCETITFGKENSTDYLYPEQLKMTVVKLISHRECOQPHYGSEV 313
 QY 696 ACLP-LWRERPQKTAISNCYITGNG-DTGRAY--SRTIQOAAIPLPKRFCEE-RYKGR-F 749
 Db 314 TTKMLCAAD-PQWK-TDSCGDSGGPLVCSLQGM-TLTGIVSGRCALCKDKPGVYTRV 370
 QY 750 TGRRLCAGNLHHRKRVSCGDSGGLMCEKRPESWVYGVTSWYGCGVADSPGVYTKV 809
 Db 371 SHPLPWIRSHTK 382
 QY 810 SAFVPIKSVTK 821

Search completed: Mon Mar 13 10:19:14 2000
 Job time : 14 secs.

THIS PAGE BLANK (USPTO)

 W E S E R F (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Run on: Mon Mar 13 10:18:23 2000; MasPar time 18.79 Seconds
 Tabular output not generated. 615.076 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (578-822) from US09147947A.pap (1 of 6)
 Perfect Score: 1892
 Sequence: 1 IIGKNSLRGMPQVSLRL.....PGVYTKVSAFVPMKSVTKL 245

Scoring table: PAM 150
 Gap 11
 Searched: 142080 seqs, 47172406 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir62
 1:pir1 2:pir2 3:pir3 4:pir4
 Statistics: Mean 44.068; Variance 74.731; scale 0.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	DB	ID	Description	Pred. No.
1752	92.6	761	2	JC5759	brain-specific serine	0.00e+00
625	33.0	625	1	KFHU1	coagulation factor XI	1.19e-116
600	31.7	638	1	KOMSP	plasma kallikrein (EC	1.09e-110
594	31.4	638	1	KORIP	plasma kallikrein (EC	2.92e-109
574	30.3	638	1	KORIP	plasma kallikrein (EC	2.92e-109
560	29.6	418	2	A34740	acrosin (EC 3.4.21.10	1.65e-104
557	29.4	437	2	A37344	acrosin (EC 3.4.21.10	3.43e-101
555	29.3	431	2	S47538	acrosin (EC 3.4.21.10	1.76e-100
547	28.9	436	2	JX0172	acrosin (EC 3.4.21.10	5.24e-100
546	28.9	559	1	A35029	acrosin (EC 3.4.21.10	4.08e-98
545	28.8	421	2	S29589	t-plasminogen activat	7.03e-98
541	28.6	431	1	UKHU	acrosin (EC 3.4.21.10	1.21e-97
542	28.6	560	1	JC4795	u-plasminogen activat	1.07e-96
536	28.3	790	1	PLPG	low-density lipoprote	6.19e-97
527	27.9	559	1	A29941	plasmin (EC 3.4.21.7)	1.61e-95
516	27.3	562	1	UKHU	t-plasminogen activat	2.14e-93
515	27.2	421	1	JC5878	t-plasminogen activat	8.30e-91
514	27.2	421	1	UKHU	acrosin (EC 3.4.21.10	2.45e-90
513	27.1	433	1	UKBA	u-plasminogen activat	7.23e-90
512	27.1	477	1	A34369	t-plasminogen activat	4.21e-90
510	27.0	455	2	A61545	plasmin (EC 3.4.21.7)	2.13e-89

hats 2 are available

24	505	26.7	655	1	A46688	hepatocyte growth fac	3.18e-88
25	504	26.6	442	1	UKPG	u-plasminogen activat	5.47e-88
26	504	26.6	460	2	B61545	plasmin (EC 3.4.21.7)	5.47e-88
27	503	26.6	812	1	PLBO	plasmin (EC 3.4.21.7)	9.38e-88
28	503	26.6	1019	1	A56318	enteropeptidase (EC 3	9.38e-88
29	501	26.5	431	2	US0599	t-plasminogen activat	2.76e-87
30	499	26.4	477	2	US0598	t-plasminogen activat	8.13e-87
31	499	26.4	477	2	US0598	t-plasminogen activat	8.13e-87
32	499	26.4	477	2	US0598	t-plasminogen activat	8.13e-87
33	489	25.8	593	2	S45281	coagulation factor XI	8.13e-87
34	489	25.8	433	1	JN0560	u-plasminogen activat	1.79e-84
35	487	25.7	810	1	PLHU	plasmin (EC 3.4.21.7)	5.25e-84
36	484	25.6	812	1	PLMS	plasmin (EC 3.4.21.7)	2.64e-83
37	481	25.4	434	1	A35005	u-plasminogen activat	1.33e-82
38	476	25.2	810	2	B30848	plasmin (EC 3.4.21.7)	1.95e-81
39	474	25.1	264	2	I38136	chymotrypsin-like pro	5.72e-81
40	474	25.1	613	2	S15468	complement C3b/C4b in	5.72e-81
41	472	24.9	433	1	UKMS	u-plasminogen activat	1.67e-80
42	472	24.9	603	2	S28941	coagulation factor XI	1.67e-80
43	470	24.8	265	2	T15451	hypothetical protein	4.90e-80
44	467	24.7	1034	1	A53653	enteropeptidase (EC 3	2.45e-79
45	463	24.5	810	2	I46260	plasmin (EC 3.4.21.7)	2.09e-78

ALIGNMENTS

```

RESULT      1
ENTRY       JC5759      #type complete
TITLE       brain-specific serine proteinase (EC 3.4.21.-) - mouse
DATE        24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change
            17-Mar-1999
ACCESSIONS  JC5759
REFERENCE   JC5759
            #authors   Yamamura, Y.; Yamashiro, K.; Tsuruoka, N.; Nakazato, H.;
            #journal    Tsujimura, A.; Yamaguchi, N.
            #title      Biochem. Biophys. Res. Commun. (1997) 239:386-392
            #comment    Molecular cloning of a novel brain-specific serine protease
            #comment    with a kringle-like structure and three scavenger receptor
            #comment    cysteine-rich motifs.
            #cross-references NID:98008848
            #accession   JC5759
            #molecule-type mRNA
            #residues    1-761 #label YAM
            ##cross-references DDBJ:D89871
            ##experimental_source brain
CLASSIFICATION #superfamily trypsin homology: scavenger receptor
                #cysteine-rich domain homology
                #glycoprotein; hydrolase; serine proteinase
FEATURES
KEYWORDS     165-157
            165-266
            166-266,273-372,
            386-486
            513-516
            517-755
            93,521,569
            562,612,711
SUMMARY      #length 761 #molecular-weight 84136 #checksum 5449
Query Match 92.6%: Score 1752; DB 2; Length 761;
Best Local Similarity 89.8%; Pred. No. 0.00e+00;
Matches 220; Conservative 17; Mismatches 8; Indels 0; Gaps 0;
Db 517 IIGKNSLRGMPQVSLRLRSASGRLCGATLLSCGVTLAAACFRGNSSYAV 576
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 578 IIGKNSLRGMPQVSLRLRSASGRLCGATLLSCGVTLAAACFRGNSSYAV 637
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 577 RGVDFTLVPEEFEDIGVOQIVIHNNYPRDSYDIALVRLQGPQCARLSTHVLPA 636
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

QY 638 RYGDYHTLVEEPEEEIGVOQIVIHREYRDRSDYDALVRLQPEEOCARFSSHVLPAC 697
Db 637 LPLMRERPKTASNCCHITGMDTGRAYSRTLQQAAPLPLPKRCKERYKGLFTGRMLCAG 696
QY 698 LPLMRERPKTASNCYITGMDTGRAYSRTLQQAAPLPLPKRCKERYKGLFTGRMLCAG 757
Db 697 NIOEDNRVDSOCCDSGGLMCEKPEDESMVYVGYTSMGCGKVTPEVTVPAFVPMIK 756
QY 758 NHEHRRVDSOCCDSGGLMCEKPEDESMVYVGYTSMGCGKVTPEVTVPAFVPMIK 817
Db 757 SVTSL 761
QY 818 SVTKL 822

RESULT 2
ENTRY #type complete
TITLE coagulation factor Xla (EC 3.4.21.27) precursor - human
SM #formal_name Homo sapiens #common_name man
#formal_name Homo sapiens #common_name man
13-Aug-1986 #sequence_revision 26-May-1994 #text_change
18-Jun-1999
ACCESSIONS A27431: A00920: A37940
REFERENCE A27431
#authors Asakai, R.; Davie, E.W.; Chung, D.W.
#journal Biochemistry (1987) 26:7221-7228
#title Organization of the gene for human factor XI.
#cross-references M0ID:88107663
#accession A27431
#molecule-type DNA
#residues 1-625 #label ASA
#cross-references GB:M18295
#note The sequence shown follows the authors' translation
REFERENCE A00920
#authors Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
#journal Biochemistry (1986) 25:2417-2424
#title Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein.
#cross-references M0ID:86243360
#accession A00920
#molecule-type mRNA
#residues 1-625 #label FUJ
#cross-references GB:M13142; NID:9182832; PIDN:AA52487.1; PID:9182833
REFERENCE A37940
#authors McMullen, B.A.; Fujikawa, K.; Davie, E.W.
#journal Biochemistry (1991) 30:2056-2060
#title Location of the disulfide bonds in human coagulation factor XI: the presence of tandem apple domains.
#cross-references M0ID:91152017
#accession A37940
#molecule-type protein
#residues 28-33:35-49, 'X', 51-55, 'X', 57-63:70-75, 'X', 77-79:107-109, 'X', 111-112:132-139, 'X', 141-154:163-164, 'X', 166-168:192, 'X', 194-198:199, 'X', 223-228:229, 'X', 231-235, 'X', 227, 248:253-254, 'X', 256-258:280-282, 'X', 284:285-287, 313-316, 'X', 318-319:320-326, 'X', 328-330, 'X', 347-349, 373, 'X', 375:377-379, 'X', 381-383:414-415, 'X', 417-431, 'X', 433-437:486-499, 'X', 501-507:535-548:559, 'X', 561-564 #label MCM

COMMENT The proenzyme consists of two identical chains linked by one or more disulfide bonds. It is activated by factor Xla (or Xli), which cleaves each chain into a light chain, which contains the active site, and a heavy chain, which associates with high molecular weight (HMW) kininogen.

GENETICS
#gene GDB:F11
#cross-references GDB:119891; OMIM:264900
#map_position 4q35-4q35
#introns 19/1: 73/2; 109/1: 162/2; 199/1: 252/2; 289/1: 343/2; 379/1: 435/2; 494/1: 526/1; 572/3

FUNCTION catalyzes the proteolytic activation of coagulation factor IX

#pathway blood coagulation intrinsic pathway
CLASSIFICATION #superfamily coagulation factor XI; trypsin homology
KEYWORDS blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydrolase; plasma; serine proteinase
FEATURE 1-18
19-387
19-108
109-198
199-288
290-379
388-625
388-618
20-103, 514-581,
571-599
29
46-76, 50-56,
110-193, 136-165,
140-146, 200-283,
226-255, 230-236,
291-374, 317-346,
321-327, 380-500,
416-432, 545-560
90, 126, 353, 450
339
387-388
431, 480, 575
491

SUMMARY #length 625 #molecular-weight 70109 #checksum 9314
Query Match 33.0%; Score 625; DB 1; Length 625;
Best Local Similarity 41.88; Pred. No. 1.19e-116;
Matches 102; Conservative 53; Mismatches 71; Indels 18; Gaps 16;

Db 388 IYGTASRGEMPMOYLH-TSPPTQ-RHLCGSIIGNOMILTAHCF--YGVESPKILR 443
QY 578 IIGKKNLRCGMPQVSLRKSSHGRLCGATLLSSCVLTAFCFRYG-NSIRSA 636
Db 444 VYSGILNQSIEKED-TSEFGVDEIITHDOYKMAESGYDALKLETVN-YT-DSQR-P 498
QY 637 VRVGD-YHTLVPEEPEEIGVOQIVIHREYRDRSDYDALVRLQPEEOCARFSSHVLP 695
Db 499 ICPLSKGDR-NVYITDQWVGWYRKLRDKIQLTOKAKIPLVINECCRRKRIYTK 557
QY 696 ACPLMRERPKTASNCYITGWS-DTGRA-YSRFTLOQAAPLPLPKRCKERYK-RTGR 752
Db 558 MICAGY-REGCK-DACKGDSGGLSC-KHNEVWHLVIGTSMGCGAQRPPGYTVVEX 614
QY 753 MLCAGNHEKRRVDSOCCDSGGLMCEKPEDESMVYVGYTSMGCGKVTPEVTVPAF 812
Db 615 VDWI 618
QY 813 VPWI 816

RESULT 3
ENTRY #type complete
TITLE plasma kallikrein (EC 3.4.21.34) precursor - mouse
ORIGIN #formal_name Mus musculus #common_name house mouse
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
18-Jun-1999
ACCESSIONS A36557
REFERENCE A36557
#authors Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachapara, L.; Rochemont, J.; Molkey, M.; Chretien, M.
#journal DNA Cell Biol. (1990) 9:737-748
#title Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison of protein and mRNA levels

COMMENT	This protein, synthesized in the liver, circulates as a noncovalent complex with high molecular weight (HMW) kininogen.	
COMMENT	The zymogen is activated by factor XIIa, which cleaves the molecule into a light chain, which contains the active site, and a heavy chain, which associates with HMW kininogen. These chains are linked by one or more disulfide bonds.	
COMMENT	The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal reaction, factor XII after its binding to a negatively charged surface. It also releases bradykinin from HMW kininogen and may also play a role in the renin-angiotensin system by converting prorenin into renin.	
GENETICS	GDB:KLUK3	
gene	#cross-references GDB:127575; OMIM:229000	
FEATURE	#map_position 4q35-4q35	
CLASSIFICATION	#superfamily coagulation factor XI; trypsin homology	
KEYWORDS	blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflammation; liver; plasma; serine protease	
20-109	#domain signal sequence #status predicted #label SIG	
110-199	#product plasma kallikrein #status predicted #label MARY	
200-289	#domain plasma kallikrein heavy chain #status predicted	
291-380	#label HGH	
391-638	#domain apple repeat #label AP1	
	#domain apple repeat #label AP2	
	#domain apple repeat #label AP3	
	#domain apple repeat #label AP4	
	#domain plasma kallikrein light chain #status predicted	
	#label LCH	
391-621	#domain trypsin homology #label TRY	
21-104,47-77,51-57,		
111-194,137-166,		
141-147,201-284,		
227-256,231-237,		
292-375,323-338,		
383-503,419-435,		
517-584,548-563,		
574-602	#disulfide_bonds #status experimental	
127,308,396,453,		
494	#binding_site carbohydrate (Asn) (covalent) #status experimental	
318-347,340-345	#disulfide_bonds #status predicted	
390-391	#cleavage_site Arg-Ile (coagulation factor XIIa) #status predicted	
	#active_site His, Asp, Ser #status predicted	
434,483,578	#length 638 #molecular_weight 71369 #checksum 585	
SUMMARY		
Query Match	31.4%	Score 594; DB 1; Length 638;
Best Local Similarity	39.6%	Pred. No. 2,92e-109;
Matches	97; Conservative	53; Mismatches 75; Indels 20; Gaps 18;
Db	391	IVGTSMSGEMWVSLQYKLT-AQ-RHLGSGSLGHQWVLTAAHCDGLDLPQWRIY 448
Y	578	ITIGKNSLRGGMWVSLRKSSHGGRLLCATLLSSCVLTAAHCKRRYG-NST-RSY 635
Db	449	SG-ILNLSDTTRKTPSQ-I-KEIITHONYKVEGSENHDAITIKIQAPLN-VTEF--OK- 500
Y	636	AVRGGDHTLTPPEE-TEELIGVQQLYIHNEYRPSDDIALVRLQSGEBOCARSSHYL 694
Db	501	PICLP-SKGDSTIYNCWVTGNGSFSEKEGIQNILOKYNIPLVYNEBCORRYODKYTO 559
Y	695	PACPLPFRERPOKTASNCYITGNG-DTGRA-YSRILQQAIPILPKRCEERYG-RETG 751
Db	560	RWACAG-KRGGR-DACKDGGSGPLCYCKNG-MWRVLGTSMSGECARREOPGYTKYAE 616
Y	752	RMICAGLHHRKRVDSQGGSGPLCEPFGSGSWVYGVTSNGVCGKVDSPGYTKYSA 811


```
RESULT 7
ENTRY A37344 #type complete
TITLE acrosin (EC 3.4.21.10) precursor form 1 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change
22-Jun-1999

ACCESSIONS A37344
REFERENCE A37344
#authors Klemm, U.; Maier, W.M.; Tsanoudou, S.; Adham, I.M.;
#journal Willison, K.; Engel, W.
#title Differentiation (1990) 42:160-166
#cross-references GB:X52466; NID:949857; PIDN:CAA36704.1; PID:949858
#accession A37344
#status preliminary
#molecule_type mRNA
#residues 1-418 #label KLE
#cross-references GB:X52466; NID:949857; PIDN:CAA36704.1; PID:949858
#classification #superfamily acrosin; trypsin homology
#note glycoprotein; hydrolase; serine proteinase

#domain trypsin homology #label TRY\
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
21-151 #disulfide_bonds #status predicted\
25-158 #disulfide_bonds #status predicted\
70-86 #disulfide_bonds #status predicted\
85,139,238 #active_site His, Asp, Ser #status predicted\
175-244 #disulfide_bonds #status predicted\
207-223 #disulfide_bonds #status predicted\
224-264 #disulfide_bonds #status predicted\
#length 418 #molecular_weight 46816 #checksum 9559

SUMMARY
Query Match 29.6%; Score 560; DB 2; Length 418;
Best Local Similarity 40.7%; Pred. No. 3,43e-101;
Matches 103; Conservative 48; Mismatches 84; Indels 18; Gaps 13;

Db 39 IVSGGSAHWGAMPWWSLQIFITSHNSRRYHACGSLNSHWVLTAAHCDFNKKRYDML 98
OY 578 IIGKNSLKGWPMQVSLRKSSHGGRLLCGATLLSCWVLTAAHCF--KRYGSTR- 633
Db 99 VFGAHEIEGRNKPVEPOERYVQKIVHEKYNVTEGNDIALKY-TPPYTCGDF--- 154
OY 634 SYAVRGDHTLVP-EFEFEIIGVQOIVIHREYRPPRSYDIALVRLOGPEOCARFSH 692
Db 155 IGPCCLPHERKAGPPRIKPHCYVTGWGCIKREAPRPSVLMEARVDLIDLCNSTQWYN 214
OY 693 VLPACLPIMRE-RPQKTASNCYITGMDTGRAYSR-T--LQQAIPLPKRFCE--ERYK 746
215 GAVTSTNVCAGY-PEGR-IDTCOGDSGGPLMCDNARQPFVYVIGTISWGVGACARARPGV 272
OY 747 GFTGRMLCAGNLHEHRRVDSGCGSGPLMC-ERPGESWVYGVYISWGVGCVKDSPGV 805
Db 273 YTAATMDYLDWIAS 285
OY 806 YTKVASAFVPIKS 818

RESULT 8
ENTRY S18407 #type complete
TITLE acrosin (EC 3.4.21.10) precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
22-Jun-1999

ACCESSIONS S18407
REFERENCE S18407
#authors Klemm, U.; Flake, A.; Engel, W.
#journal Biochim. Biophys. Acta (1991) 1090:270-272
#title Rat sperm acrosin: cDNA sequence, derived primary structure
and phylogenetic origin.
#cross-references MUID:92031708
```

```
#accession S18407
#molecule_type mRNA
#residues 1-437 #label KLE
#cross-references EMBL:X59254
REFERENCE S30037
#authors Klemm, U.; Flake, A.; Engel, W.
#submission submitted to the EMBL Data Library, April 1991
#accession S30037
#molecule_type mRNA
#residues 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 #label KLE2
#cross-references EMBL:X59254; NID:957282; PIDN:CAA11947.1; PID:957283
REFERENCE A56620
#authors Kremling, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
#journal DNA Seq. (1991) 2:57-60
#title Exon-intron structure and nucleotide sequence of the rat
proacrosin gene.
#cross-references MUID:92199245
#accession A56620
#status preliminary
#molecule_type DNA; mRNA
#residues 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 #label KRE
#note sequence modified after extraction from NCBI backbone
sequence extracted from NCBI backbone (NCBIN:89436,
NCBIN:89439, NCBIN:89447, NCBIN:89454, NCBIN:89458)
#superfamily acrosin; trypsin homology
glycoprotein; hydrolase; serine proteinase; sperm; zymogen
FEATURE
KEYWORDS 1-19
20-43 #domain signal sequence #status predicted #label SIG\
43-286 #domain acrosin light chain #status predicted #label
44-437 #domain acrosin heavy chain #status predicted #label
HCH\
22,211 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
75-155,29-163 #disulfide_bonds #status predicted\
74-90 #disulfide_bonds #status predicted\
89,143,241 #active_site His, Asp, Ser #status predicted\
178-247 #disulfide_bonds #status predicted\
210-226 #disulfide_bonds #status predicted\
237-267 #disulfide_bonds #status predicted\
#length 437 #molecular_weight 48279 #checksum 1220

SUMMARY
Query Match 29.4%; Score 557; DB 2; Length 437;
Best Local Similarity 40.9%; Pred. No. 1.76e-100;
Matches 103; Conservative 46; Mismatches 86; Indels 17; Gaps 11;

Db 43 IVGGGSSRRAMPWWSLQIFITSHNSRRYHACGSLNSHWVLTAAHCDFNKKRYDML 102
OY 578 IIGKNSLKGWPMQVSLRKSSHGGRLLCGATLLSCWVLTAAHCF--KRYGSTR- 633
Db 103 VFGAHEIEGRNKPVEPOERYVQKIVHEKYNVTEGNDIALKY-TPPYTCGDF--- 158
OY 693 VLPACLPIMRE-RPQKTASNCYITGMDTGRAYSR-T--LQQAIPLPKRFCE--ERYK 747
215 GAVTSTNVCAGY-PEGR-IDTCOGDSGGPLMCDNARQPFVYVIGTISWGVGACARARPGV 276
OY 748 GFTGRMLCAGNLHEHRRVDSGCGSGPLMC-ERPGESWVYGVYISWGVGCVKDSPGV 806
Db 277 YTAATMDYLDWIAS 288
OY 807 YTKVASAFVPIKS 818

RESULT 9
ENTRY S47538 #type complete
TITLE acrosin (EC 3.4.21.10) precursor - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
```


REFERENCE	AJ159/
#authors	Ny, T.; Leonardsson, G.; Hsueh, A.J.W.

REFERENCE	A31597
#authors	
	Ny, T.; Leonardsson, G.; Hsueh, A.J.W.

#journal DNA (1988) 7:671-677
#title Cloning and characterization of a cDNA for rat tissue-type
#cross-references M01D:89170114
#accession A31597
#molecule-type mRNA
##residues 1-379, 'K', 381-559 #label NYT
##cross-references GB:M2697; NID:9530159; PIDN:AAA1812.1; PID:9530160
CLASSIFICATION #superfamily tissue plasminogen activator; EGF homology;
#fibrinectin type I repeat homology; kringle homology;
#trypsin homology
#fibrinolysis; glycoprotein; hydrolase; kringle; serine
protease
KEYWORDS
FEATURE
1-17 #domain signal sequence #status predicted #label SIG
18-29 #domain propeptide #status predicted #label PRO
30-559 #product t-plasminogen activator #status predicted
#label MAT
30-308 #product t-plasminogen activator chain A #status
#predicted #label ACH
30-75 #domain fibronectin type I repeat homology #label 1F1
30-116 #domain EGF homology #label EGF
124-205 #domain kringle homology #label KR1
213-294 #domain kringle homology #label KR2
309-559 #product t-plasminogen activator chain B #status
#predicted #label BCH
309-553 #domain trypsin homology #label TRY
38-68,66-75,83-94,
88-105,107-116,
124-205,145-187,
176-200,213-284,
234-276,265-289,
297-428,340-356,
348-417,442-516,
474-490,506-534
149,481
#disulfide_bonds #status predicted
#binding_site carbohydrate (asn) (covalent) #status
predicted
308-309 #cleavage_site Arg-Ile (plasmin, trypsin) #status
predicted
#active_site His, Asp, Ser #status predicted
355,404,510 #length 559 #molecular-weight 62903 #checksum 8573
SUMMARY
Query Match 28.8%; Score 545; DB 1; Length 559;
Best Local Similarity 38.3%; Pred. No. 7.03e-98;
Matches 93; Conservative 58; Mismatches 76; Indels 16; Gaps 13;
Db 321 PQOAFVKKRSPGFEFLGGVLISSCWVSAHCEVERF--PPHHLKVLGRTYR-VV 377
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
590 PMOVSRLKSSHGOG-KLLCGATLLSSCWVITAAHCF-KRYGNSTRSAVRYG-DYHILV 646
QY 378 PGEEDQTEIEKYIVHKEFDDDTYNDIALIQLRSDSSQCAOESSVGTACLPD-PDYOL 436
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
647 PEEFEIEIGVOQIVIHREYRPRSDYDIALVRLGPEPQCARFSHVLPACLPMLREHPQ 706
QY 437 PMTECELSGYKHEASSPFSDRLKEAHVLYSSRCTSOHLNKTITSMNLACAGDRT 496
:
707 KTASNCYITITGWMG--DTRRAY-SRTLOQAIPLPKRRC-EER-YKGRPTGMLCAGNLH- 760
QY 497 GGNODVHADCGSGGGLVC-MIDKRMTLIGISMGAGCGKDVPGITTKTYNLINMTQD 555
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
761 -EKRRV-DSOGDSGGPLMCEPESWVYGVTSMTGICGVKDSBGTITKVSAPFWIKS 818
QY 556 NMK 558
QY 819 VTK 821
RESULT 12
ENTRY S29599 #type fragment
TITLE acrosin (EC 3.4.21.10) precursor - guinea pig (fragment)
ORGANISM #formal name Cavia porcellus - common name guinea pig
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
22-Jun-1999

ACCESSIONS S29599
REFERENCE S29599
#authors Gerton, G.L.; Hoff, H.B.; Baba, T.
#submitter submitted to the EMBL Data Library, May 1992
#description The amino acid sequence of guinea pig proacrosin deduced from
its cDNA sequence.
#accession S29599
#molecule-type mRNA
##residues 1-421 #label GER
##cross-references EMBL:212153; NID:949559; PIDN:CAA78137.1; PID:949560
CLASSIFICATION #superfamily acrosin; trypsin homology
KEYWORDS glycoprotein; hydrolase; serine proteinase
FEATURE
41-264 #domain trypsin homology #label TRY
SUMMARY #length 421 #checksum 9563
Query Match 28.8%; Score 545; DB 2; Length 421;
Best Local Similarity 39.1%; Pred. No. 1.21e-97;
Matches 99; Conservative 46; Mismatches 89; Indels 19; Gaps 14;
Db 41 IIGGQTAPGAMPVMSLQIFMAHNNRRYHACGILLNSHWVLTAAHCFDSKRYDWRL 100
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 578 IIGGKNSLRGMPQVSLRLKSSHGDRLL-LCGATLLSSCWVITAAHCF--KRYGNSTR- 633
QY 101 VFGAELEIFYGNNKRYRPLQERY-VEKIVTHEKNTYNEGNDIALKLTTPPV-CGPF-- 156
:
QY 634 SYAVRVGDY--HTLVPEEIEIGVOQIVIHREYRPRSDYDIALVRLGPEPQCARFS 691
157 -IGPGCLPTFRAGPKIPQICGYVAGWYIREKAPRPSVLLEAVELIDDLGNSQWYN 215
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 692 HVLACLPMLRERQKTAISNCTITGMDT--GRA-Y-SRTLOQAIPLPKRFE--ERYK 746
QY 216 GRVNSTVYACGY-BEGR-IDTCGDSGGLMCRDANSPFVWYGISWVGACARARPGI 273
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 747 GRFGRLMCAQNLHHRKRVKSCGDSGGLMC-ERPESWVYGVTSMTGICGVKDSBGPV 805
Db 274 YTAWDVYIDWIAS 286
QY 806 YTKVASAPFWIKS 818
RESULT 13
ENTRY UKH1 #type complete
TITLE u-plasminogen activator (EC 3.4.21.73) precursor - human
ALTERNATE_NAMES cellular plasminogen activator; urokinase-type
plasminogen activator (uPA)
CONTAINS urokinase-type plasminogen activator chain A; urokinase-type
plasminogen activator chain A1; urokinase-type plasminogen
activator chain B; urokinase-type plasminogen activator,
single chain form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-Dec-1982 #sequence_revision 04-Dec-1986 #text_change
18-Jun-1999
ACCESSIONS A00931; I52209; J0102; A37561; I38102; S65783; A37562;
A37563; A37564; A35689; A36097
REFERENCE A00931
#authors Riccio, A.; Grimaldi, G.; Verde, P.; Sebastio, G.; Boast, S.;
Blasi, F.
#journal Nucleic Acids Res. (1985) 13:2759-2771
#title The human urokinase-plasminogen activator gene and its
promoter.
#cross-references M01D:85215647
#accession A00931
##molecule-type DNA
##residues 1-431 #label RIC
##cross-references GB:X02419; NID:937601; PIDN:CAA26268.1; PID:e300604;
PID:g1834524
##note the authors translated the codon ATG for residue 214 as
Ile
REFERENCE I52209
#authors Nagamine, Y.; Pearson, D.; Gratian, M.
#journal Biochem. Biophys. Res. Commun. (1985) 132:563-569
#title Exon-Intron boundary sliding in the generation of two mRNAs

coding for porcine urokinase-like plasminogen activator.

#cross-references MUID:86050639

#accession 152209

#status preliminary; translated from GB/EMBL/DBJ

#molecule-type DNA

#residues 145-161 ##label NAG1

#cross-references GB:K03027; NID:g330174; PIDN:AAA61257.1; PID:g340175

REFERENCE

#authors Nagai, M.; Hiramatsu, R.; Kaneda, T.; Hayasuke, N.; Arimura, H.; Nishida, M.; Suyama, T.

#journal Gene (1985) 36:183-188

#title Molecular cloning of cDNA coding for human preprourokinase.

#cross-references MUID:86056554

#accession J70102

#molecule-type mRNA

#residues 1-213,'I',215-431 ##label NAG2

#cross-references GB:K03226; NID:g340155; PIDN:AAC97138.1; PID:g340158; GB:D00244; NID:g220138; PID:d1000623; PID:g220139

REFERENCE

#authors A37561

Verde, P.; Stoppelli, M.P.; Galeffi, P.; Di Nocera, P.; Blasi, F.

#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:4727-4731

#title Identification and primary sequence of an unspliced human urokinase poly(A)+ RNA.

#cross-references MUID:84272706

#accession A37561

#molecule-type mRNA

#residues 66-431 ##label VER

#cross-references GB:D00244; NID:g220138

REFERENCE

#authors 138102

Jacobs, P.; Cravador, A.; Loriau, R.; Brockly, F.; Colau, B.; Chuchana, P.; van Elsen, A.; Herzog, A.; Bollen, A.

#journal DNA (1985) 4:139-146

#title Molecular cloning, sequencing, and expression in *Escherichia coli* of human preprourokinase cDNA.

#cross-references MUID:85203359

#accession 138102

#status preliminary

#molecule-type mRNA

#residues 1-150,'W',152-213,'I',215-385,'C',387-429,'V',431

##label JAC

##cross-references EMBL:X02760; NID:g35297; PIDN:CAA26535.1; PID:g35298

REFERENCE

#authors 565783

Yoshimoto, M.; Ushiyama, Y.; Sakai, M.; Tamaki, S.; Hara, H.; Takahashi, K.; Sawasaki, Y.; Hanada, K.

#journal Biochim. Biophys. Acta (1996) 1293:83-89

#title Characterization of single chain urokinase-type plasminogen activator with a novel amino-acid substitution in the kringle structure.

#cross-references MUID:96186279

#accession S65783

#status preliminary

#molecule-type mRNA

#residues 21-140,'L',142-213,'I',215-431 ##label YOS

#cross-references EMBL:D11143; NID:g1311467; PIDN:BAA01919.1; PID:d1002396; PID:g1199928

REFERENCE

#authors A37562

Gunzler, W.A.; Steffens, G.J.; Otting, F.; Kim, S.M.A.; Franks, E.; Flohe, L.

#journal Hoppe-Seyler's Z. Physiol. Chem. (1982) 363:1155-1165

#title The primary structure of high molecular mass urokinase from human urine.

#cross-references MUID:83055084

#accession A37562

#molecule-type protein

#residues 21-177 ##label GUN

REFERENCE

#authors A37563

Schaller, J.; Nick, H.; Rickli, E.E.; Gillesen, D.; Lergier, W.; Studer, R.O.

#journal Eur. J. Biochem. (1982) 125:251-257

#title Human low-molecular-weight urinary urokinase. Partial characterization and preliminary sequence data of the two polypeptide chains.

#cross-references MUID:83003608

#accession A37563

#molecule-type protein

#residues 156-176;179-193,'T',195,'T',197-224 ##label SCH

REFERENCE

#authors A37564

Steffens, G.J.; Gunzler, W.A.; Otting, F.; Franks, E.; Flohe, L.

#journal Hoppe-Seyler's Z. Physiol. Chem. (1982) 363:1043-1058

#title The complete amino acid sequence of low molecular mass urokinase from human urine.

#cross-references MUID:83055099

#accession A37564

#molecule-type protein

#residues 158-410 ##label STE

REFERENCE

#authors A35689

Kentzer, E.J.; Buko, A.; Menon, G.; Sarin, V.K.

#journal Biochem. Biophys. Res. Commun. (1990) 171:401-406

#title Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase.

#cross-references MUID:90365737

#accession A35689

#molecule-type protein

#residues 21-30,'X',32,'X',34-38,'X',40-43 ##label KEN

##note Identification of a fucose and attempt to determine its attachment site

REFERENCE

#authors A36697

Rabbani, S.A.; Desjardins, J.; Bell, A.W.; Banville, D.; Mazur, A.; Henkin, J.; Goltzman, D.

#journal Biochem. Biophys. Res. Commun. (1990) 173:1058-1064

#title An amino-terminal fragment of urokinase isolated from a prostate cancer cell line (PC-3) is mitogenic for osteoblast-like cells.

#cross-references MUID:91097529

#accession A36697

#molecule-type protein

#residues 21-34 ##label RAB

REFERENCE

#authors A51255

Li, X.; Bokman, A.M.; Llinas, M.; Smith, R.A.G.; Dobson, C.M.

#submission submitted to the Brookhaven Protein Data Bank, July 1993

#cross-references PDB:1KDU

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR, residues 69-153

REFERENCE

#authors A44375

Li, X.; Smith, R.A.G.; Dobson, C.M.

#journal Biochemistry (1992) 31:9562-9571

#title Sequential (1)H NMR assignments and secondary structure of the kringle domain from urokinase.

#cross-references MUID:93003110

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR

REFERENCE

#authors A66822

Hansen, A.P.; Petros, A.M.; Meadows, R.P.; Nettesheim, D.G.; Mazur, A.P.; Olejniczak, E.T.; Xu, R.X.; Pederson, T.M.; Henkin, J.; Resik, S.W.

#submission submitted to the Brookhaven Protein Data Bank, January 1994

#cross-references PDB:1U9K

#contents annotation; conformation and disulfide bond assignments by (1)H-NMR, residues 26-155

REFERENCE

#authors A66058

Spraggan, G.S.; Phillips, C.; Nowak, U.K.; Ponting, C.P.; Saunders, D.; Dobson, C.M.; Stuart, D.I.; Jones, E.Y.

#submission submitted to the Brookhaven Protein Data Bank, July 1995

#cross-references PDB:1LWM

#contents annotation; X-ray crystallography, 2.5 angstroms, residues 168-175;179-426

COMMENT This enzyme is found in urine in a high molecular mass form, consisting of A and B chains, and a low molecular mass form, consisting of A1 and B chains.

COMMENT Urokinase-type plasminogen activator proteolytically activates plasminogen, and the inactive single-chain form is proteolytically activated by plasmin (see PR:PLHU).

GENETICS

#gene GDB:PLAU


```

#accession      heart.
#status         JE0315
#molecule_type preliminary
#residues       1-1113 #label TOM
##cros-references DBRef:AB013874
CLASSIFICATION #superfamily trypsin homology
FEATURE
SUMMARY 869-1097 #domain trypsin homology #label TRY
          length 1113 #molecular-weight 122984 #checksum 224

Query Match      28.6%; Score 542; DB 2; Length 1113;
Best Local Similarity 40.7%; Pred. No. 6,19e-97;
Matches 98; Conservative 39; Mismatches 90; Indels 14; Gaps 11;

869 ILGRTSRPGKMPWQCSIQ--SEPS-GHI-CGCVLIARKWVLTVAHCFEGREDAD-VWKV 923
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
578 ILGKNSLRGMPWQVSLRLKSSHGDRLLCGATLLSSCWVLTAAHCFKRYGNSTRSYAV 637
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 924 VEGINNLDPGSGEMQTRFVKITLLHPRYSRVAVDYDISVEL--SDI-NE-TSYVRPVC 979
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 638 RVGDYHTLVPEEFEEIGVOQIVIHREYRPDRSDYDALVRLQPEEQCAFSSSHVLPAC 697
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 980 LP-SPEEYLEPDTYCYITGMGHMGNKMPFKLQEGEVRIIPLEQCQSYFDMKTTNRMICA 1038
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 698 LPLWREPRQKTASNCYITGMGTGRAYSRLLQQAIPLLPKRFCEERYKGR-FTGRMLCA 756
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1039 G--YESGTVDSMGDSGGPLVCERPGGQWTLFGLTSMGVCFSKVLPGVYSNYSYFVGW 1096
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 757 GNLHEHKKRVDSGCGSGPLMCEKRGESWVYGVTSWYGCGVKD-SFGYITKVSALFVPM 815
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1097 I 1097
QY 816 I 816

```

Search completed: Mon Mar 13 10:18:43 2000
 Job time : 20 secs.

THIS PAGE BLANK (USPTO)

227-327/5ROR2

Use bit No. 1

(TM)

Distribution rights by Oxford Molecular Ltd

t generated

1 IRLAGGSSVHEGRV

PAM 15C

131253 seqs, 129

Listing first 45

a-issued

mean 26.593; V=

ved by analysis

0.4 1168 1 C45 76


```

CC INFORMATION FOR SEQ ID NO: 2
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 451 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;
SQ
CC
CC Query Match 56.5%; Score 428; DB 2; Length 451;
CC Best Local Similarity 53.5%; Pred. No. 7,26e-34;
CC Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
Dy 350 VRLVGGSPHNGRREILHLSQMGWICDDREVRNGVYVCSLGPVQVAHKAHNGCT 409
Qy 227 IRLAGGSVHEGRVELYHAQMGVCCDDWDADAEVICHQLGLSGIAKAMHQAATGEGS 286
Dy 410 GPIMLNEVFCGFRSSIEECKIRQMGTRACSHSEDAVCTCT 450
Qy 287 GPVMLDEYRCTGNELSLIECPKSSWGHNCGHKEDAGVCT 327
Xx
Xx RESULT 6 STANDARD: PRT; 451 AA.
Xx ID US-08-154-365-2
Xx AC xxxxxx
Xx DT
Xx DE
Xx Sequence 2, Application US/08154365
Xx
Xx Sequence 2, Application US/08154365
Xx Patent No. 5624904
Xx GENERAL INFORMATION:
Xx APPLICANT: Dunne, Dana W.
Xx APPLICANT: Resnick, David
Xx APPLICANT: Kreiger, Monty
Xx APPLICANT: Joiner, Keith A.
Xx TITLE OF INVENTION: Method for Treating Gram-Positive
Xx TITLE OF INVENTION: Septicemia
Xx NUMBER OF SEQUENCES: 2
Xx CORRESPONDENCE ADDRESS:
Xx ADDRESSEE: Patrea L. Pabst
Xx STREET: 1100 Peachtree Street, Suite 2800
Xx City: Atlanta
Xx STATE: Ga
Xx COUNTRY: USA
Xx ZIP: 30309-4530
Xx
Xx COMPUTER READABLE FORM:
Xx MEDIUM TYPE: Floppy disk
Xx COMPUTER: IBM PC compatible
Xx OPERATING SYSTEM: PC-DOS/MS-DOS
Xx SOFTWARE: Patentin Release #1.0, Version #1.25
Xx CURRENT APPLICATION DATA:
Xx APPLICATION NUMBER: US/08/154,365
Xx FILING DATE:
Xx CLASSIFICATION: 514
Xx ATTORNEY/AGENT INFORMATION:
Xx NAME: Pabst, Patrea L.
Xx REGISTRATION NUMBER: 31,284
Xx REFERENCE/DOCKET NUMBER: MIT6392
Xx TELECOMMUNICATION INFORMATION:
Xx TELEPHONE: (404)-815-6508
Xx TELEFAX: (404)-815-6555
Xx INFORMATION FOR SEQ ID NO: 2:
Xx SEQUENCE CHARACTERISTICS:
Xx LENGTH: 451 amino acids
Xx TYPE: amino acid
Xx STRANDEDNESS: single
Xx TOPOLOGY: linear
Xx MOLECULE TYPE: protein
Xx HYPOTHETICAL: YES
Xx ANTI-SENSE: NO
Xx FRAGMENT TYPE: N-terminal

```

```
CC ORIGINAL SOURCE:  
CC ORGANISM: homo sapien  
CC PUBLICATION INFORMATION:  
CC AUTHORS: Ashkenas, et al.  
CC JOURNAL: J. Lipid Res.  
CC VOLUME: 34  
CC PAGES: 983-1000  
CC DATE: 1993  
CC RELEVANT RESIDUES IN SEQ ID NO: 1 : FROM 1 TO 451  
SO SEQUENCE 451 AA; 49841 MM; 102052 CN;  
  
Query Match 56.4%; Score 427; DB 1; Length 451;  
Best Local Similarity 53.5%; Pred. No. 9,24e-34;  
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;  
  
Db 350 VLVLGSSPHGEKREVIILHSGMGITCDNNMEYRVGGVCRSLGPRGOAVHKAAHFQGT 409  
227 IRLAGSSSVHGSRVELYHAAGMGTVCDDQMDADAELVICRQLGLGIKAHWQAYFEGGS 286  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
410 GPWLNVEVFCFGRESSTIECKIRQMKTRACSHSDAGVTCT 450  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
287 GPWLDEVRTGNELSTIECPKRSSWGERHNCGHKEADGVST 327  
QY  
  
RESULT 7 STANDARD: PRT; 585 AA.  
ID US-08-473-791-10  
XX xxxxxx  
XX  
XX  
XX  
XX  
XX Sequence 10, Application US/08473791  
DE  
XX  
Sequence 10, Application US/08473791  
CC Patent No. 5736340  
CC GENERAL INFORMATION:  
CC APPLICANT: Koths, Kirston E.  
CC APPLICANT: Halenbeck, Robert F.  
CC APPLICANT: Taylor, Eric W.  
CC APPLICANT: Wang, Alice M.  
CC APPLICANT: Caspipt, Clayton L.  
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein  
CC NUMBER OF SEQUENCES: 11  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Celus Oncology Corporation  
CC STREET: 1400 Fifty-Third street  
CC CITY: Emeryville  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94608  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/473,791  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/316,714  
CC FILING DATE:  
CC APPLICATION NUMBER: US/07/961,404  
CC FILING DATE: 15-OCT-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Goldman, Kenneth M.  
CC REGISTRATION NUMBER: 34,174  
CC REFERENCE/DOCKET NUMBER: 2595.1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 420-3152  
CC TELEFAX: (510) 658-5470  
CC TELEX: N/A  
CC INFORMATION FOR SEQ ID NO: 10:
```

```
CC      SEQUENCE CHARACTERISTICS:  
CC      LENGTH: 585 amino acids  
CC      TYPE: amino acid  
CC      STRANDEDNESS: Single  
CC      TOPOLOGY: linear  
CC      MOLECULE TYPE: protein  
CC      SEQUENCE 585 AA; 65330 MW; 1.874129 CN;  
SQ  
  
Query Match          54.4%;   SCORE 412; DB 1; Length 585;  
Best Local Similarity 53.5%;   Fred.No. 3,44e-32;  
Matches    54; Conservative 19; Mismatches 28; Indels 0; Gaps 0.  
  
Db      24 MRADGATNGGREIFEFIRGOMGTVCNDLMLDTASVYCRALGFENATQALGRAAFSGS 83  
Oy      227 IRLAGSSVHGGRVELHLAGONGIVCDQDDMDADAELVICRQLGLSIGAKAHQAIFEGGS 286  
Db      84 GPIMLDEVQCTGTETASLADCKSLGWLSNCRHERDAGVCT 124  
Oy      287 GPVMLEVRCTGNELSTEQCPKSSWGEHNCHKKEDAGVCT 327  
  
RESULT      8  
ID US-08-316-714-10 STANDARD; PRT; 585 AA.  
XX xxxxxx  
XX  
XX  
DE Sequence 10, Application US/08316714  
XX  
CC Sequence 10, Application US/08316714  
CC Patent No. 5965382  
CC GENERAL INFORMATION:  
CC APPLICANT: Kochs, Kirston E.  
CC APPLICANT: Halenbeck, Robert F.  
CC APPLICANT: Taylor, Eric W.  
CC APPLICANT: Wang, Alice M.  
CC APPLICANT: Casipit, Clayton L.  
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein  
CC NUMBER OF SEQUENCES: 11  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSSEE: Cetus Oncology Corporation  
CC STREET: 1400 Fifty-Third Street  
CC CITY: Emeryville  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94608  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/316,714  
CC  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/961,404  
CC FILING DATE: 15-OCT-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Goldman, Kenneth M.  
CC REGISTRATION NUMBER: 34,174  
CC REFERENCE/DOCKET NUMBER: 2595.1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (510) 420-3152  
CC TELEFAX: (510) 658-5470  
CC  
CC TELEX: N/A  
CC INFORMATION FOR SEO ID NO: 10:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 585 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC
```

CC MOLECULE TYPE: Protein
SQ SEQUENCE 585 AA; 65330 MW; 1874129 CN;
Query Match 54.4%; Score 412; DB 2; Length 585;
Best Local Similarity 53.5%; Pred. No. 3,44e-32;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;
DB 24 MRIADGAINOGVEVEFYRGOMGTVCNMDLMDLSDASVCRALGFENATQALGRAAFGGGS 83
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQMDADAEVICRQLGSLGIAKAMHQAIFYEGGS 286
DB 84 GPIMDEVOCCTGEASLADCKSLGWLKSNCRHERDAGVCT 124
QY 287 GPVLMDEVRCCTGNELSTIEQCPKSSWGEHNGHKEDAGVSCT 327
T 9
US-08-477-674-10 STANDARD; PRT: 585 AA.
AC xxxxxx
DE Sequence 10, Application US/08477674
XX
XX
XX
CC Sequence 10, Application US/08477674
CC Patent No. 5644035
CC GENERAL INFORMATION:
CC APPLICANT: Kolhs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casibit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,674
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2395.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 585 AA; 65330 MW; 1874129 CN;






Query Match 54.4%; Score 412; DB 1; Length 585;
Best Local Similarity 53.5%; Pred. No. 3,44e-32;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;
DB 24 MRIADGAINOGVEVEFYRGOMGTVCNMDLMDLSDASVCRALGFENATQALGRAAFGGGS 83
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQMDADAEVICRQLGSLGIAKAMHQAIFYEGGS 286
DB 84 GPIMDEVOCCTGEASLADCKSLGWLKSNCRHERDAGVCT 124
QY 287 GPVLMDEVRCCTGNELSTIEQCPKSSWGEHNGHKEDAGVSCT 327
RESULT 10
ID US-08-470-350B-2 STANDARD; PRT: 1290 AA.
AC xxxxxx
DE Sequence 2, Application US/08470350B
XX
XX
XX
CC Sequence 2, Application US/08470350B
CC Patent No. 5684126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H
CC TITLE OF INVENTION: Edoerin: A Secreted von Ebner's Gland
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107,48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1290 AA; 141879 MW; 9315616 CN;
Query Match 48.2%; Score 365; DB 1; Length 1290;
Best Local Similarity 52.4%; Pred. No. 2,75e-27;
Matches 54; Conservative 17; Mismatches 28; Indels 4; Gaps 4;
DB 292 VRLVNGDRGRCGRVVELLYQ-GSMGTVCDDSDWTKDANYVCRL-VCGWALSAPGSANHEQ 349
QY 227 IRLAGSSVHEGRVE-LYHAGOMGTVCDDQMDADAEVICRQLGSLG-IKAMHQAIFYEGE 284
DB 350 GSGSTVLDDVACTGHEAYILMSCSHRGWLSHNGCHHEDAGVTS 392
QY 285 GSGPVLMDEVRCCTGNELSTIEQCPKSSWGEHNGHKEDAGVSCT 327

RESULT 11
ID US-08-794-795-7 STANDARD: PRT: 489 AA.
AC xxxxxx
XX
XX
DT
XX
Sequence 7, Application US/08794795
XX
Sequence 7, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Eishourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794.795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 489 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 489 AA; 49441 MW; 1141296 CN;
SQ
Query Match 46.9%; Score 355; DB 2; Length 489;
Best Local Similarity 50.5%; Pired. No. 3.00e-26;
Matches 51; Conservative 23; Mismatches 22; Indels 5; Gaps 5;
Db 394 VRIMGITN-R-GRAEYNNENGTICDDDDNNNDATVFCMIGYS-RGRA-LSY-GGGS 448
QY 227 IRLAGSSVHEGRVELYHAGQMGTCDDDDADAEVICRQLGSLGIAKAMHQAIFEGGS 286
Db 449 GNIMLDVNCRGTEINSLMDCSKNSMGHNHCYVHEDAGVECS 489
QY 287 GPVMDDEVCTGNETLSIEQPKSSWGEHNCGRHEDAGVCT 327
RESULT 12
ID US-08-392-367B-2 STANDARD: PRT: 518 AA.
AC xxxxxx
XX
XX
DT
XX

XX
DT
XX
DE
XX
XX
Sequence 2, Application US/08392367B
CC Patent No. 5691197
CC GENERAL INFORMATION:
CC APPLICANT: Tryggvason, Karl
CC APPLICANT: Eliomaa, Outi
CC APPLICANT: Kangas, Maarit
CC TITLE OF INVENTION: An Insolated DNA Sequence For a
CC Patent No. 5691197
CC TITLE OF INVENTION: No. 5691197a1 Macrophage Receptor with
CC TITLE OF INVENTION: a Collagenous Domain and the
CC TITLE OF INVENTION: Polypeptide Chain Encoded by
CC TITLE OF INVENTION: such a Sequence
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Fay, Sharpe, Beall, Fagan,
CC ADDRESSEE: Minnich & McKee
CC STREET: 1100 Superior Avenue
CC CITY: Cleveland
CC STATE: Ohio
CC COUNTRY: U.S.A.
CC ZIP: 44114-2518
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch,
CC MEDIUM TYPE: 720 Kb storable
CC COMPUTER: IBM PS/2, Model 35 SX
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: Word Perfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/392.367B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Minnich, Richard J.
CC REGISTRATION NUMBER: 24,175
CC REFERENCE/DOCKET NUMBER: TRY 2 009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (216) 861-5582
CC TELEFAX: (216) 241-1666
CC TELEX: (216) 980162
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 518 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: Single
CC TOPOLOGY: Linear
CC SEQUENCE 518 AA; 52730 MW; 1281000 CN;
SQ
Query Match 46.9%; Score 355; DB 1; Length 518;
Best Local Similarity 50.5%; Pired. No. 3.00e-26;
Matches 51; Conservative 23; Mismatches 22; Indels 5; Gaps 5;
Db 423 VRIMGITN-R-GRAEYNNENGTICDDDDNNNDATVFCMIGYS-RGRA-LSY-GGGS 477
QY 227 IRLAGSSVHEGRVELYHAGQMGTCDDDDADAEVICRQLGSLGIAKAMHQAIFEGGS 286
Db 478 GNIMLDVNCRGTEINSLMDCSKNSMGHNHCYVHEDAGVECS 518
QY 287 GPVMDDEVCTGNETLSIEQPKSSWGEHNCGRHEDAGVCT 327
RESULT 13
ID US-08-794-795-2 STANDARD: PRT: 495 AA.
AC xxxxxx
XX
XX
DT
XX

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eplor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows, Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA: 49764 MW: 1218428 CN;
SO
Query Match 44.6%; Score 338; DB 2; Length 495;
Best Local Similarity 46.5%; Pred. No. 1,72e-24;
Matches 47; Conservative 25; Mismatches 24; Indels 5; Gaps 4;
Db 399 VRIVGSSN-R-GRAEYVYSTGTICDDDEQNSDAIVFCMIGYS-KGRALYKV--GAGT 453
QY 227 IRLAGSSVHEGVELYHAGVGWTCDDDDADAEVIGICROLGSLAKAMHQAIFYGEGS 286
Db 454 GOIWLNVOCRGTESTLWSTCKNSMGHDCSHEDAGVCS 494
QY 287 GPVMDDEVRCGTGNELSLIEQCPKSSMGHNGCHREDAGVSC 327
RESULT 14
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:

CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eplor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA: 52658 MW: 1340662 CN;
SO
Query Match 44.6%; Score 338; DB 2; Length 520;
Best Local Similarity 46.5%; Pred. No. 1,72e-24;
Matches 47; Conservative 25; Mismatches 24; Indels 5; Gaps 4;
Db 424 VRIVGSSN-R-GRAEYVYSTGTICDDDEQNSDAIVFCMIGYS-KGRALYKV--GAGT 478
QY 227 IRLAGSSVHEGVELYHAGVGWTCDDDDADAEVIGICROLGSLAKAMHQAIFYGEGS 286
Db 479 GOIWLNVOCRGTESTLWSTCKNSMGHDCSHEDAGVCS 519
QY 287 GPVMDDEVRCGTGNELSLIEQCPKSSMGHNGCHREDAGVSC 327
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08200900A
XX
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

1. pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:28:17 2000; MasPar time 10.34 Seconds
460.695 Million cell updates/sec
Tabular output not generated.

```

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLGGSSVHEGRPELYHAG.....KSSWGHNCGHKEDAGVSC 101

Scoring table: PAM 150

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database: pir62

Statistics: Mean 37.625; Variance 64.653; scale 0.582

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	length	DB	ID	Description	Pred. No.
1	704	93.0	761	2	JC5759	brain-specific serine	1.86e-13
2	444	58.7	453	2	S08276	LDL receptor I, macro	6.47e-77
3	436	57.6	454	2	I46862	macrophage scavenger	4.66e-75
4	430	56.8	458	2	B4407	macrophage scavenger	1.15e-73
5	428	56.5	451	2	A38415	macrophage scavenger	3.33e-73
6	428	56.5	454	2	A4407	macrophage scavenger	3.33e-73
7	427	56.4	2153	2	T14893	scavenger receptor, cy	5.68e-73
8	416	55.0	1116	2	S36077	M130 antigen - human	1.99e-70
9	416	55.0	1149	2	I38006	M130 antigen (extracel	1.99e-70
10	416	55.0	1151	2	I38004	M130 antigen (cytosol	1.99e-70
11	416	55.0	1156	2	I38005	M130 antigen (cytosol	1.99e-70
12	412	54.4	585	2	A47161	Mec-2-binding glycopr	1.67e-69
13	401	53.0	504	2	S56745	mucin (clone pGM3-1)	5.73e-67
14	397	52.4	918	2	JC4361	scavenger receptor, cy	4.77e-66
15	384	50.7	577	2	A55202	cyclophilin C-associa	4.61e-66
16	384	50.3	600	2	S55704	mucin (clone pGM7-1)	2.24e-62
17	373	49.3	1436	2	A46496	antigen WCL1, precurs	1.52e-60
18	365	48.2	1290	2	A57190	epherin precursor - r	1.02e-58
19	355	46.9	518	2	A55840	macrophage bacteria-b	1.92e-56
20	325	42.9	468	2	S26741	T-cell glycoprotein C	1.17e-49
21	321	42.4	626	2	I49100	mec6f precursor - mou	9.25e-49
22	293	38.7	532	2	A37251	spectact receptor prec	1.67e-42
23	158	20.9	1034	1	A35665	enteropeptidase (EC 3	6.17e-14

use bit: No. 2

24	127	16.8	495	1	●A51270	T-cell surface glycop	4.72e-08
25	126	16.6	1019	1	A56318	entropепidase (EC 3	7.19e-08
26	125	16.5	1035	1	A43090	entropепidase (EC 3	1.09e-07
27	115	16.5				g-protein coupled r	6.66e-06
28	115	15.2	442	2	T47074	T-cell surface glycop	2.22e-05
29	112	14.8	495	1	A26396	complement C3b/Cb in	5.15e-04
30	104	13.7	613	2	S15468	hypothetical protein	7.56e-04
31	103	13.6	403	2	T02513	T-cell surface glycop	1.62e-03
32	101	13.3	501	1	A43535	lymphocyte surface gl	3.46e-03
33	99	13.1	494	1	A29079	restriction modifict	1.53e-02
34	95	12.5	425	2	B64316	hypothetical protein	6.55e-02
35	91	12.0	607	2	B70920	hypothetical protein	9.37e-02
36	87	11.5	226	2	D69156	conserved hypotetica	2.70e-01
37	86	11.4	810	2	S67050	probable membrane pro	3.83e-01
38	85	11.4	1113	2	JEO317	low-density lipoprote	3.83e-01
39	85	11.2	768	2	S76837	pled-4 protein - Sme	5.40e-01
40	84	11.1	76	2	T39960	hypothetical protein	7.61e-01
41	84	11.1	558	2	T15448	cellulase (EC 3.2.	7.61e-01
42	84	11.1	741	2	A47063	Doc protein, stress-	7.61e-01
43	84	11.0	2825	2	T14271	vacuolar processing e	1.07e-00
44	83	11.0	494	2	T05372	furin (EC 3.4.21.75)	1.07e-00
45	83	11.0	1680	2	A43434		

ALIGNMENTS

RESULT	ENTRY	1
TITLE	JC5759	#type complete
ORGANISM		brain-specific serine proteinase (EC 3.4.21.-) - mouse
DATE		#format_name Mus musculus #common_name house mouse
		24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 17-Mar-1999
ACCESSIONS	JC5759	
REFERENCE	JC5759	
#authors		Yamamura, Y.; Yamashiro, K.; Tsunoka, N.; Nakazato, H.;
#journal		Tsujimura, A.; Yamaguchi, N.
#title		Biochem. Biophys. Res. Commun. (1997) 239:386-392
		Molecular cloning of a novel brain-specific serine protease with a kringle-like structure and three scavenger receptor cysteine-rich motifs.
#cross-references		MTID:9800848

```

#cross-references MVID:98008848
accession JC5759
##molecule_type mRNA
##residues 1-761 ##label YAM
##cross-references DDBJ:D89871
##experimental_source brain
CLASSIFICATION #superfamily trypsin homology; scavenger receptor
cysteine-rich domain homology
KEYWORDS glycoprotein; hydrolase; serine proteinase
FEATURE

```

```

166-266,273-372,
386-486
#domain scavenger receptor cysteine-rich #status
#predicted #label SRC\
513-516 #domain furin binding #status predicted #label FRB\
517-755 #domain trypsin homology #label TRY\
93,521,569 #binding site carbonylrate (asn) (covalent) #status
#predicted\
562,612,711 #active site His, Asp, Ser #status predicted
SUMMARY #length 761 #molecular weight 84136 #checksum 5449

```

Query Match	93.08;	Score 704;	DB 2;	Length 761;
Best Local Similarity	90.08;	Pred. No. 1.86e-138;		
Matches	90;	Conservative	7;	Mismatches 3; Indels 0; Gaps 0;

[illegible]

[illegible]

```

FEATURE                                     cysteine-rich domain homology
350-453                                  #domain scavenger receptor cysteine-rich domain homology
                                           #label SRC
SUMMARY                                #length 454 #molecular-weight 49745 #checksum 2903
Query Match                               57.6%; ● Score 436; DB 2; Length 454;
Best Local Similarity 54.5%: Pred. No. 4,656-75;
Matches 55; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

Db      353 VRLVGGRPHEGRVEYLHNGQWGTCCDHWELRAGGVCRSLGYRGVSVHKRAYFGGCT 412
       ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Oy      227 IRLAGSSVHEGRVELYHAGQMGTCDDQDDADAEVICROLGLSGIAKAMHQAYFEGGS 286
       ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db      413 GPVLNVEPCLGMSSIEECKIKONGVRVCSSHEDAGVICT 453
       ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Oy      287 GPVMLDEVRCETGNELSTEOCPKSSWGEBHNCKREDAGVSCT 327

RESULT          4
ENTRY           B44407                #type complete
TITLE           macrophage scavenger receptor, MSRI - mouse
ALTERNATE_NAMES macrophage scavenger receptor type I
ORGANISM        #format_name Mus musculus #common_name house mouse
DATE            30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
               20-Sep-1999
ACCESSIONS     B44407; A38260; I56334
REFERENCE
#authors       Doi, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.;
               Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.;
               Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
#journal       J. Biol. Chem. (1993) 268:2126-2133
#title         Charged collagen structure mediates the recognition of
               negatively charged macromolecules by macrophage scavenger
               receptors.
#cross-references MIMD:93131972
#accession     B44407
#status        preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues      1-458 #label DOI
#experimental_source macrophage-like cell line P38BD
#note          sequence extracted from NCBI backbone (NCBIP:123207)
REFERENCE
#authors       Freeman, M.; Ashkenas, J.; Rees, D.J.G.; Kingsley, D.M.;
               Copeland, N.G.; Jenkins, N.A.; Krieger, M.
               Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8810-8814
#journal       An ancient, highly conserved family of cysteine-rich protein
               domains revealed by cloning type I and type II murine
               macrophage scavenger receptors.
#cross-references MIMD:91062370
#accession     A38260
#status        preliminary
#molecule_type mRNA
#residues      349-458 #label FRE
#cross-references GB:M59445; GB:M36817; NID:9192737; PIDN:AAA37464.1;
               PID:9192738
REFERENCE
#authors       Ashkenas, J.; Penman, M.; Vasile, E.; Acton, S.; Freeman,
               M.W.; Krieger, M.
               J. Lipid Res. (1993) 34:983-1000
#journal       Structures and high and low affinity ligand binding
               properties of murine type I and type II macrophage
               scavenger receptors.
#cross-references MIMD:93359822
#accession     I56334
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues      5-458 #label RES
#cross-references GB:L04274; NID:g293745; PIDN:AAA39747.1; PID:g293746
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
               cysteine-rich domain homology
FEATURE
354-457                                  #domain scavenger receptor cysteine-rich domain homology

```


[illegible]

Query Match

Query Match 56.48; Score 427; DB 2; Length 2153;

Best Local Similarity 55.4%; Pred. No. 5,686-73;
Matches 56; Conservative 23; Mismatches 22; Indels 0; Gaps 0;

Db 462 VRLVGLNREGREVEIFLNNMGTCVDDMGTPDANVYCRQLGPGSGARSARFEGGS 521
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 522 VPLLDNVCSSGERSLELCSNNGIGVHNGCHDSDASVCT 562
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 287 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 327
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

RESULT 8
ENTRY S36077 #type complete
TITLE M130 antigen - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change

IONS 138003; S36077
NCE 138003
Ctors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal #title
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

CLASSIFICATION #molecule_type mRNA
#residues 1-116 #label RES
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
FEATURE #superfamily scavenger receptor cysteine-rich domain homology

43-147 #domain scavenger receptor cysteine-rich domain homology.
151-254 #label SRC1\
#domain scavenger receptor cysteine-rich domain homology
258-361 #label SRC2\
#domain scavenger receptor cysteine-rich domain homology
365-468 #label SRC3\
#domain scavenger receptor cysteine-rich domain homology
470-573 #label SRC4\
#domain scavenger receptor cysteine-rich domain homology
575-678 #label SRC5\
#domain scavenger receptor cysteine-rich domain homology
711-814 #label SRC6\
#domain scavenger receptor cysteine-rich domain homology
816-920 #label SRC7\
#domain scavenger receptor cysteine-rich domain homology
-1024 #label SRC8\
#domain scavenger receptor cysteine-rich domain homology
Y #label SRC9
#domain scavenger receptor cysteine-rich domain homology

Query Match 55.0%; Score 416; DB 2; Length 1116;
Best Local Similarity 55.4%; Pred. No. 1,996-70;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

Db 924 IRLQEGPTSCGRVEIWHGSMGTVCDDSDLDADAQVCCQLGCGPALKFAKFAEFGGT 983
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 984 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 1024
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 287 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 327
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

RESULT 9
ENTRY 138006 #type complete
TITLE M130 antigen (extracellular variant) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change

ACCESSIONS 138006; S35768

REFERENCE 138003
#authors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 924 IRLQEGPTSCGRVEIWHGSMGTVCDDSDLDADAQVCCQLGCGPALKFAKFAEFGGT 983
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 984 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 1024
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 287 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 327
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

RESULT 10
ENTRY 138004 #type complete
TITLE M130 antigen (cytosolic variant 1) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change

ACCESSIONS 138004; S36078
REFERENCE 138003
#authors Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 1017 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 1057
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 287 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 327
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

SUMMARY #length 1149 #molecular-weight 124328 #checksum 487

Query Match 55.0%; Score 416; DB 2; Length 1149;
Best Local Similarity 55.4%; Pred. No. 1,996-70;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

Db 957 IRLQEGPTSCGRVEIWHGSMGTVCDDSDLDADAQVCCQLGCGPALKFAKFAEFGGT 1016
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 227 IRLAGSSVHEGRVELYHAGMGTCVDDQWDADAIEICRLGLSGIAKAMHQAIFEGGS 286
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

Db 1017 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 1057
#journal #title
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
#superfamily scavenger receptor cysteine-rich domain homology

Qy 287 GPVMDDEVCTGNELISIEQPKSSWGEHNCGRKEDAGVCT 327
#cross-references M01D:93380506
#status preliminary; translated from GB/EMBL/DBJ

CLASSIFICATION #molecule_type mRNA
#residues 1-1151 #label RES
#cross-references EMBL:Z22968; NID:9312147; PID:9312148
FEATURE #superfamily scavenger receptor cysteine-rich domain homology

43-147 #domain scavenger receptor cysteine-rich domain homology
151-254 #label SRC1\
#domain scavenger receptor cysteine-rich domain homology
258-361 #label SRC2\
#domain scavenger receptor cysteine-rich domain homology
365-468 #label SRC3\
#domain scavenger receptor cysteine-rich domain homology
470-573 #label SRC4\
#domain scavenger receptor cysteine-rich domain homology
608-711 #label SRC5\
#domain scavenger receptor cysteine-rich domain homology
744-847 #label SRC6\
#domain scavenger receptor cysteine-rich domain homology
849-953 #label SRC7\
#domain scavenger receptor cysteine-rich domain homology
954-1057 #label SRC8\
#domain scavenger receptor cysteine-rich domain homology
#label SRC9


```
##molecule_type mRNA
##residues 1-577 ##label CHI
##cross-references EMBL:X67809; NID:G297032; PID:G297033; GB:X67803
##note sequence appears consistent with either a GPI anchor or
```

REFERENCE

A48231 transmembrane domain near the carboxyl end

#authors

Friedman, J.; Trahey, M.; Weissman, I.

#journal

Proc. Natl. Acad. Sci. U.S.A. (1993) 90:6815-6819

#title

Cloning and characterization of cyclophilin C-associated protein: a candidate natural cellular ligand for cyclophilin C.

#cross-references W01D:9342080

#accession A48231

#molecule-type mRNA

#residues 1-24, 'G', '26-227', 'P', '229-465', 'NE', '468-572', 'LH' ##label

GENETICS ##cross-references GB:L16894; NID:9397799; PID:9397800

CLASSIFICATION

CycAP

#superfamily scavenger receptor cysteine-rich domain homology cell surface component; glycoprotein; membrane protein

#gene

#residues 1-124

#domain signal sequence #status predicted #label SIG\ #domain scavenger receptor cysteine-rich domain homology #label SRC

SUMMARY #length 577 #molecular-weight 64491 #checksum 6506

Query Match 50.7%; Score 384; DB 2; Length 577;

Best Local Similarity 50.5%; Pred. No. 4,61e-63; Mismatches 15; Indels 0; Gaps 0;

Db 24 MRLVNGASANEGRVEIFYRGRMGTVCDNLWMLDAHVCRALGYENATQALGRAAFGPGK 83

QY 227 IRLAGSSVHEGRVELYHAGQWGVCDQWDADAEYICROLGLSGTAKAMHQAIFEGS 286

Db 84 GPIMLDEVECTGTESSLASCRSLGMYVRCGHEKDAVCS 124

QY 287 GPVWLDEVCTGNELSIQCPSKSSWGEHNCGRKEDAGVCT 327

Search completed: Mon Mar 13 10:28:30 2000
Job time: 13 secs.

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Nucleotide sequence (TM)

h-p protein - protein database search, using Smith-Waterman algorithm
Mon Mar 13 10:27:15 2000; Maspar time 6.46 Seconds
466,945 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLAGSSVHGKVELYHAG.....KSMGEHNGHKEDAGVSC 101

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 38.669; Variance 61.361; scale 0.630

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	875	1	NETR_HUMAN NEUTROTYPIN PRECURSOR	2.71e-163
2	708	93.5	761	1	NETR_MOUSE NEUTROTYPIN PRECURSOR	1.46e-150
3	444	58.7	453	1	MSRE_BOVIN MACROPHAGE SCAVENGER R	4.98e-83
4	436	57.6	454	1	MSRE_RABIT MACROPHAGE SCAVENGER R	5.04e-81
5	430	56.8	458	1	MSRE_MOUSE MACROPHAGE SCAVENGER R	1.60e-79
6	428	56.5	451	1	MSRE_HUMAN MACROPHAGE SCAVENGER R	5.07e-79
7	373	49.3	1436	1	MC1L_BOVIN ANTIGEN MC1.1	2.33e-65
8	325	42.9	468	1	CD6_HUMAN T-CELL DIFFERENTIATION	1.30e-53
9	293	38.7	532	1	CD6_STRUP EGG PEPTIDE SPECTR RE	6.69e-46
10	158	20.9	1034	1	ENTK_PIG ENTEROPEPTIDASE PRECUR	7.27e-09
11	127	16.8	495	1	CD5_BOVIN T-CELL SURFACE GLYCOPR	1.14e-08
12	126	16.6	1019	1	ENTK_HUMAN ENTEROPEPTIDASE PRECUR	1.78e-08
13	125	16.5	1035	1	ENTK_BOVIN ENTEROPEPTIDASE PRECUR	1.63e-07
14	120	15.9	1069	1	ENTK_MOUSE ENTEROPEPTIDASE PRECUR	1.63e-07
15	112	14.8	495	1	CD5_HUMAN T-CELL SURFACE GLYCOPR	5.18e-06
16	101	13.3	491	1	CD5_RAT T-CELL SURFACE GLYCOPR	4.99e-04
17	99	13.1	494	1	CD5_MOUSE T-CELL SURFACE GLYCOPR	5.42e-03
18	92	12.5	343	1	Y130_METJA HYPOTHETICAL PROTEIN M	1.73e-02
19	92	12.2	492	1	TMS2_HUMAN TRANSMEMBRANE PROTEIN	2.53e-02
20	91	12.0	603	1	SYT_SYNY3 THREONYL-TRNA SYNTHETA	3.39e-01
21	84	11.1	741	1	GUNS_CLOM ENDOGLUCANASE SS PRECU	4.86e-01
22	83	11.0	1680	1	FUR2_DROME FURIN-LIKE PROTEASE 2	9.86e-01
23	81	10.7	212	1	YM95_MYCTU HYPOTHETICAL 23.3 KD P	9.86e-01

RESULT	1	STANDARD	PRT	875 AA.	ALIGNMENTS
ID	NETR_HUMAN				
AC	P66730;				
DT	15-DEC-1999 (Rel. 39, Crea)				
DT	15-DEC-1999 (Rel. 39, Last sequence update)				
DT	15-DEC-1999 (Rel. 39, Last annotation update)				
DE	NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).				
GN	PRSS12				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-BRAIN:				
RX	MEDLINE; 98201705.				
RA	PROBA K., GSCHWEND T.P., SONDERGER P.;				
RT	Cloning and sequencing of the cDNA encoding human neutrotypsin.;				
RL	Biochim. Biophys. Acta 1396:143-147(1998).				
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.				
CC	*****				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: A001531; CA004816.1; -				
DR	PROSITE: PS00134; TRYPSIN_HIS; 1.				
DR	PROSITE: PS00135; TRYPSIN_SER; 1.				
DR	PROSITE: PS00420; SPERACT_RECEPTOR; 3.				
KW	Hydrolyase; serine protease; Glycoprotein; Kringle; Repeat; Signal.				
FT	CHAIN	1	20	POTENTIAL.	
FT	CHAIN	21	875	NEUTROTYPIN.	
FT	DOMAIN	23	92	PROLINE-RICH.	
FT	DOMAIN	93	165	KRINGLE.	
FT	DOMAIN	170	271	SRCR 1.	
FT	DOMAIN	280	381	SRCR 2.	
FT	DOMAIN	387	487	SRCR 3.	

[illegible]

Query Match	Best Local Match	Similarity	Score	DB	Length
Matches	91: Conservative	93.5%	708	1	761
				Pred. No. 1,46e-150;	
				Mismatches 3;	
				Indels 0;	
				Gaps 0;	

Query	Match	Similarity	Score	DB	Length
Db	166	IRLVGNSGHEGRVLYHAGQ	761 AA; 84118 MW; 3FC3CA	F35 CRC32;	
Qy	227	IRLGGSSVHEGRVLYHAGQ	761 AA; 84118 MW; 3FC3CA	F35 CRC32;	
Db	226	GPVHDEVRCTGNELSTIEQ	761 AA; 84118 MW; 3FC3CA	F35 CRC32;	
Qy	287	GPVHDEVRCTGNELSTIEQ	761 AA; 84118 MW; 3FC3CA	F35 CRC32;	

RESULT	3	STANDARD;	PRT;	453 AA.
ID	MSRE BOVIN			
AC	P21758;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED			
DE	IDL RECEPTOR I AND II).			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;			
CC	Bovinae; Bos.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE..			
RC	TISSUE=LUNG;			
RX	MEDLINE: 90136965.			
RA	KODAMA T., FREEMAN M., ROHRER L., ZABECKY J., MATSUDAIRA P.,			
RA	KRIEGER M.;			
RT	"Type I macrophage scavenger receptor contains alpha-helical and			
RT	collagen-like coiled coils.";			
RL	Nature 343:531-535(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A. (SHORT FORM).			
RC	TISSUE=LUNG;			
RX	MEDLINE: 90136973.			
RA	ROHRER L., FREEMAN M., KODAMA T., PENNAN M., KRIEGER M.;			
RT	"Coiled-coil fibrous domains mediate ligand binding by macrophage			
RT	scavenger receptor type II.";			
RL	Nature 343:570-572(1990).			
CC	-I- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC			
CC	DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROSCLEROSIS.			
CC	TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE			
CC	ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING			


```
CC CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC -1- SUBUNIT: HOMOTRIMER.
CC CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
EMBL; X51689; CA035987.1; -.
EMBL; X54183; CA038108.1; -.
PIR; S08276; S08276.
PIR; S08278; S08278.
PROSITE; PS00420; SPERACT_RECEPTOR_1.
DR PFAM; PF00530; SRCR; 1.
KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; IDL; Heptad repeat pattern; Receptor; Alternative splicing.
FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
FT TRANSHEM 51 76 SIGNAL ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 77 108 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 109 271 SPACER (PROBABLE).
FT DOMAIN 272 343 ALPHA-HELICAL COILED-COIL.
FT DOMAIN 344 453 COLLAGEN-LIKE.
FT CARBOHYD 82 82 CYS-RICH.
FT CARBOHYD 101 101 POTENTIAL.
FT CARBOHYD 142 142 POTENTIAL.
FT CARBOHYD 183 183 POTENTIAL.
FT CARBOHYD 220 220 POTENTIAL.
FT CARBOHYD 248 248 POTENTIAL.
FT CARBOHYD 266 266 POTENTIAL.
FT VASAPLIC 348 349 QS -> PG (IN ISOFORM II).
FT VASAPLIC 350 453 MISSING (IN ISOFORM II).
SQ SEQUENCE 453 AA; 50056 MW; 2CDEIBEI CRC32;

Query Match 58.7%; Score 444; DB 1; Length 453;
Best Local Similarity 55.4%; Pred. No. 4,98e+83;
Matches 56; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

      352 VRLVSGSPHGGRVYFIPEHGGMGTVCDDDRHWELRGGLGVCRSLGYGVSVMKRAAFEGKT 411
          ||| |||| |||||| :||||||| |: ||| || | : :||| | :
      227 IRLAGSSVHEQRLLHHAGOMGTVCDDQMDADAEVICRQLGSIGAKMHQAQFAFGES 286
          ||: |::| ::|||::| ||: |||::| || : |::|::| ||
      OY 287 GPVMLDEVRCGNELSTIEDCPRSSGMGHNCGRKEADAGVSC T 327


RESULT 4
ID MSRE_RABIT STANDARD; PRT; 454 AA.
AC Q05585;
DT 01-OCT-1996 (Rel. 34, Created)
DI 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED LDL RECEPTOR I AND II).
DE MSR1.
CS Cryotlagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93016877.
RA BICKEL P.E., FREEMAN M.W.: "Rabbit aortic smooth muscle cells express inducible macrophage scavenger receptor messenger RNA that is absent from endothelial
```

```

RT      cells."
RL      J. Clin. Invest. 90:1450-1457(1992).
[2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 93131972.
RA      DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
RA      MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KANABE Y.,
RA      ITRAKURA H., YAZAKI Y., MATSUMOTO A.;
RT      "Charged collagen structure mediates the recognition of negatively
RT      charged macromolecules by macrophage scavenger receptors."
J. Biol. Chem. 268:2126-2133(1993).
CC      -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
CC      DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC      TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
CC      ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
CC      MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC      -1- SUBUNIT: HOMOTRIMER.
CC      -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC      -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
CC      RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
CC      OF THE SAME GENE.
-----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; L11693; AAA31402.1; -
DR      EMBL; L11692; AAA31403.1; -
DR      EMBL; D13381; BAA02649.1; -
DR      PROSITE; PS00420; SPERACT_RECEPTOR; 1.
DR      PFM; PF00391; SRCR; 1.
KW      Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
KW      Heptad repeat pattern; Receptor; Alternative splicing.
FT      DOMAIN 1 50
FT      TRANSMEM 51 73
FT      FT
FT      DOMAIN 74 454
FT      DOMAIN 74 109
FT      DOMAIN 110 272
FT      DOMAIN 273 344
FT      DOMAIN 345 454
FT      DOMAIN 82 82
FT      CARBOHYD 102 102
FT      CARBOHYD 143 143
FT      CARBOHYD 184 184
FT      CARBOHYD 221 221
FT      CARBOHYD 249 249
FT      CARBOHYD 267 267
FT      VARSPPLIC 348 354
FT      VARSPPLIC 355 454
FT      CONFLICT 106 106
SQ      SEQUENCE 454 AA; 49745 MW; 5D780348 CRC32;
Query Match 57.6%; Score 436; DB 1; Length 454;
Best Local Similarity 34.5%; Fred. No. 5.04e-81;
Matches 55; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
Dh 353 VRLVGGRGPHGREGREILHNGOMGTIVCDHMLRAGOVVCRSLGTYGVKSVHKRAYFGOGT 412
Oy 227 IRLAGSSVHGGRVELYHAGOMGTIVCDQDMDDALAEVICTRLGLSGSLAKANHQATFEGS 286
Dh 413 GPIWLNEVPCLGMESSIEECKIRQMGVAVGVCSHGDAVGT 453
Oy 287 GPVMLDEVRTGNTLSIEQCPKSSWGHNGCHKEDAGVSGT 327

```


CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PFM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC
CC EMBL: U09860; AAC50138.1; -
CC HSSP: P00763; IDPO.
CC MIM: 226200; -
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS00180; CUB; 2.
CC PROSITE: PS00740; MAM_1; 1.
CC PROSITE: PS00060; MAM_2; 1.
CC PROSITE: PS01209; LDLRA_1; 2.
CC PROSITE: PS50068; LDLRA_2; 2.
CC PFM: PF00057; 1d1_recept_a; 2.
CC PFM: PF00089; trypsin; 1.
CC PFM: PF00431; CUB; 2.
CC PFM: PF00530; SRCR; 1.
CC PFM: PF00629; MAM; 1.
CC PFM: PF01390; SEA; 1.
CC Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
CC Serine protease; Zymogen; Transmembrane; Repeat.
CC CHAIN 1 784
CC TRANSMEM 19 47
CC DOMAIN 182 223
CC DOMAIN 225 334
CC DOMAIN 342 504
CC DOMAIN 524 634
CC DOMAIN 641 679
CC DOMAIN 678 771
CC ACT_SITE 825 835
CC ACT_SITE 876 876
CC ACT_SITE 971 971
CC LIPID 2 2
CC DISULFID 184 197
CC DISULFID 191 210
CC DISULFID 204 221
CC DISULFID 643 655
CC DISULFID 650 668
CC DISULFID 662 677
CC DISULFID 772 836
CC DISULFID 810 826
CC DISULFID 910 977
CC DISULFID 941 956
CC DISULFID 967 995
CC CARBOHYD 116 116
CC CARBOHYD 147 147

FT CARBOHYD 179 179 POTENTIAL.
FT CARBOHYD 328 328 POTENTIAL.
FT CARBOHYD 335 335 POTENTIAL.
FT CARBOHYD 388 388 POTENTIAL.
FT CARBOHYD 440 440 POTENTIAL.
FT CARBOHYD 470 470 POTENTIAL.
FT CARBOHYD 503 503 POTENTIAL.
FT CARBOHYD 534 534 POTENTIAL.
FT CARBOHYD 630 630 POTENTIAL.
FT CARBOHYD 682 682 POTENTIAL.
FT CARBOHYD 706 706 POTENTIAL.
FT CARBOHYD 725 725 POTENTIAL.
FT CARBOHYD 848 848 POTENTIAL.
FT CARBOHYD 887 887 POTENTIAL.
FT CARBOHYD 909 909 POTENTIAL.
FT CARBOHYD 949 949 POTENTIAL.
SQ SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;
Query Match 16.68; Score 126; DB 1; Length 1019;
Best Local Similarity 27.18; Pred. No. 1,14e-08;
Matches 26; Conservative 23; Mismatches 41; Indels 6; Gaps 6;
Db 678 VREFNGTNNGLVRRIRIOSIWHITACAEKMTQISNDVCGILGL-GSGNSKPIFTDG- 735
QY 227 IRLAGSSVHEGRVLYHAGQMTVCDDWDADAEVTCRQJLSIAKAWHQAIFGSS 286
Db 736 GPEVKINTAP-DGHLITPSQCLDPSLIRLOCNHR 770
QY 287 GP-VMLDEVNCTGN-ELS-IEQPKSSWGEHNGHK 319
RESULT 13
ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-DUODENUM;
RX MEDLINE: 94329561.
RA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 94043122.
RA LAVALLE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
RA FERENZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE-INTESTINE;
RX MEDLINE: 92189715.
RA LIGHT A., JANSKA H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN

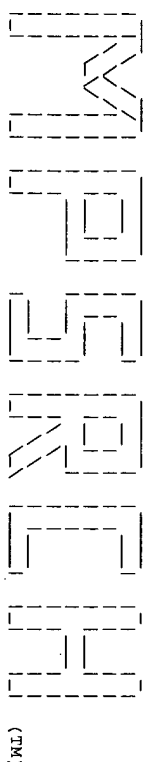
CC TRYPsinogen.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPsin-LIKE PROTEASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U09859; ABA0026.1; -
 CC EMBL: L19663; AAA1035.1; -
 CC PIR: A61436; A61436.
 CC HSP: P00763; IDPO.
 CC PROSITE: PS00134; TRYPsin_HIS. 1.
 CC PROSITE: PS00135; TRYPsin_SER. 1.
 CC PROSITE: PS01180; CUB. 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS50060; MAM_2; 1.
 CC PROSITE: PS01209; LDLRA_1; 2.
 CC PROSITE: PS50068; LDLRA_2; 2.
 CC PFAM: PF00057; Idl_receptor_a; 2.
 CC PFAM: PF00089; trypsin; 1.
 CC PFAM: PF00431; CUB; 2.
 CC PFAM: PF00530; SRCR; 1.
 CC PFAM: PF00629; MAM; 1.
 CC PFAM: PF01390; SEA; 1.
 CC Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
 CC Setine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 CC CHAIN 1 800
 CC CHAIN 801 1035
 CC TRANSMEM 19 47
 CC DOMAIN 197 238
 CC DOMAIN 240 350
 CC DOMAIN 358 520
 CC DOMAIN 540 650
 CC DOMAIN 657 695
 CC DOMAIN 694 787
 CC DOMAIN 841 841
 CC ACT_SITE 892
 CC ACT_SITE 897
 CC LIPID 2
 CC DISULFID 199 212
 CC DISULFID 206 225
 CC DISULFID 219 236
 CC DISULFID 659 671
 CC DISULFID 666 684
 CC DISULFID 678 693
 CC DISULFID 788 812
 CC DISULFID 826 842
 CC DISULFID 926 933
 CC DISULFID 957 972
 CC DISULFID 983 1011
 CC CARBOHYD 116
 CC CARBOHYD 147
 CC CARBOHYD 170
 CC CARBOHYD 194
 CC CARBOHYD 233
 CC CARBOHYD 263
 CC CARBOHYD 264
 CC CARBOHYD 404

FT CARBOHYD 456 456 POTENTIAL.
 FT CARBOHYD 486 486 POTENTIAL.
 FT CARBOHYD 519 519 POTENTIAL.
 FT CARBOHYD 550 550 POTENTIAL.
 FT CARBOHYD 646 646 POTENTIAL.
 FT CARBOHYD 698 698 POTENTIAL.
 FT CARBOHYD 722 722 POTENTIAL.
 FT CARBOHYD 741 741 POTENTIAL.
 FT CARBOHYD 762 762 POTENTIAL.
 FT CARBOHYD 864 864 POTENTIAL.
 FT CARBOHYD 903 903 POTENTIAL.
 FT CARBOHYD 965 965 POTENTIAL.
 FT VARSPIC 166 192 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 808 808 R -> Y (IN REF. 3).
 SQ SEQUENCE 1035 AA; 114887 MW; 5173034A CRC32;
 Query Match 16.5%; Score 125; DB 1; Length 1035;
 Best Local Similarity 26.8%; Pred. No. 1,78e-08;
 Matches 22; Conservative 21; Mismatches 35; Indels 4; Gaps 4;
 Db 694 VRLNGTDSGLVQFRIOSIMHYACAENMTTQISDDVCLLGL-GTGNSVPT-FSTGG 751
 QY 227 IRLAGSSVHGRIYELHACQMGVCCDDADAEVTCRLGLSIKAMHQAIFEGS 286
 Db 752 GPVYNTAPNGSLILPPSOQC 773
 QY 287 GP-VMLDEVRCCTGWEIS-IEQC 306
 RESULT 14
 ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=DUDENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.;
 Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPsin AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPsinogen TO TRYPsin WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPsinogen,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPsinogen.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPsin-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U73378; AAB37317.1; -
 DR HSSP: P00763; IDPO.
 DR MGD: MG1:1197523; PRSS7.
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 DR PROSITE: PS01180; CUB. 2.
 DR PROSITE: PS00740; MAM. 1; 1.
 DR PROSITE: PS50060; MAM. 2; 1.
 DR PROSITE: PS01209; LDLR. 1; 2.
 DR PROSITE: PS50068; LDLR. 2; 2.
 DR PFAM: PF00057; 1dl_recept. a; 2.
 DR PFAM: PF00089; trypsin. 1.
 DR PFAM: PF00431; CUB. 2.
 DR PFAM: PF00350; SRCR. 1.
 DR PFAM: PF00629; MAM. 1.
 DR PFAM: PF01390; SEA. 1.
 DR Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 Serine protease; Zymogen; Transmembrane; Repeat;
 CHAIN 1 829
 TRANSMEM 19 1069
 DOMAIN 227 268
 FT 270 379
 FT 387 549
 FT 569 679
 FT 686 724
 FT 723 816
 FT 874 874
 FT 925 925
 FT ACT SITE 1021 1021
 FT LIPID 2 2
 FT DISULFID 229 242
 FT DISULFID 236 255
 FT DISULFID 249 266
 FT DISULFID 688 700
 FT DISULFID 695 713
 FT DISULFID 707 722
 FT DISULFID 817 945
 FT DISULFID 859 875
 FT DISULFID 959 1027
 FT DISULFID 991 1006
 FT DISULFID 1017 1045
 FT CARBOHYD 147 147
 FT CARBOHYD 197 197
 FT CARBOHYD 212 212
 FT CARBOHYD 373 373
 FT CARBOHYD 380 380
 FT CARBOHYD 433 433
 FT CARBOHYD 515 515
 FT CARBOHYD 579 579
 FT CARBOHYD 675 675
 FT CARBOHYD 727 727
 FT CARBOHYD 751 751
 FT CARBOHYD 770 770
 FT CARBOHYD 791 791
 FT CARBOHYD 897 897
 FT CARBOHYD 936 936
 FT CARBOHYD 999 999
 SO SEQUENCE 1069 AA; 118735 MW; 48B825A2 CRC32;
 Query Match 15.98; Score 120; DB 1; Length 1069;
 Best Local Similarity 25.38; Pred. No. 1.63e-07;
 Matches 24; Conservative 24; Mismatches 43; Indels 4; Gaps 4;
 Db 723 VRLNGRSNGIYQFMHSHIACAEENWTQISNEVCHLIGT-GSANS-SMPISNGG 780
 QY 227 IRLAGSSVHEGRVELYHAQGMGTVCDDQMDADAEVLCROGLSGIAKAWHQAIFEGGS 286
 Db 781 GPEVRVNAQNGSLILTPSLQCSODSLILLOCNHR 815
 QY 287 GP-VMLDEVACTGNELSIE-QCPKSSWGEHNGCHK 319

RESULT 15
 ID CD5_HUMAN STANDARD; PRT; 495 AA.
 AC P06127;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN
 DE T1/LEU-1) (LYMPHOCYTE ANTIGEN CD5).
 GN CD5 OR LEU1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87014786.
 RA JONES N.H., CLABBY M.L., DIALYNAS D.P., HUANG H.-J.S.,
 RA HERZENBERG L.A., STROMINGER J.L.;
 RT "Isolation of complementary DNA clones encoding the human lymphocyte
 RT glycoprotein T1/Leu-1";
 RL Nature 323:346-349(1986).
 RN [2]
 RP INTERACTION WITH CD72/LYB-2.
 RX MEDLINE; 91270374.
 RA VAN DE VELDE H., VON HOEGEN I., LHO W., PARNES J.R., THIELEMAN K.;
 RT "The B-cell surface protein CD72/Lyb-2 is the ligand for CD5.";
 RL Nature 351:662-665(1991).
 CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL
 CC -1- SUBCELLULAR LOCATION: CD5 INTERACTS WITH CD72/LYB-2.
 CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd5.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04391; CAA27979.1; -
 DR PIR: A26396; A26396.
 DR MIM: 153340; -
 DR PFAM: PF00530; SRCR. 3.
 KW Signal; Transmembrane; Glycoprotein; T-cell.
 FT SIGNAL 1 24
 FT CHAIN 25 495
 FT DOMAIN 25 372
 FT TRANSMEM 373 402
 FT DOMAIN 403 495
 FT DOMAIN 34 133
 FT DOMAIN 275 382
 FT DISULFID 44 107
 FT CARBOHYD 116 116
 FT CARBOHYD 241 241
 SO SEQUENCE 495 AA; 54625 MW; 8507C34C CRC32;
 Query Match 14.88; Score 112; DB 1; Length 495;
 Best Local Similarity 40.58; Pred. No. 5.18e-06;
 Matches 17; Conservative 11; Mismatches 12; Indels 2; Gaps 1;
 Db 277 RLVGSSICEGTEVRGAQMAALCDSSARSRLRWEVCRE 318
 QY 228 RLAGSSVHEGRVELYHAQGMGTVCDDQMDADA--EVICRQ 267
 Search completed: Mon Mar 13 10:27:25 2000
 Job time: 10 secs.

227-324/SPCR2
Asc Int No.3



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Protein - protein database search, using Smith-Waterman algorithm
on: Mon Mar 13 10:27:42 2000; Maspar time 16.64 Seconds
420.878 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (227-327) from US09147947A.pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLAGSSVHEGRVELYHAG.....KSSWGEHNCGHKEDAGVSC 101

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mnc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 37.892; Variance 61.971; scale 0.611

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	450	59.4	2043	5	096943	SRCR DOMAIN, MEMBRANE	4.97e-82
2	427	56.4	2153	5	097375	SCAVENGER RECEPTOR CYS	2.00e-76
3	416	55.0	1116	4	007898	M130 ANTIGEN PRECURSOR	9.37e-74
4	416	55.0	1149	4	007901	M130 ANTIGEN, EXTRACEL	9.37e-74
5	416	55.0	1151	4	007899	M130 ANTIGEN, CYTOPLAS	9.37e-74
6	416	55.0	1156	4	007900	M130 ANTIGEN, CYTOPLAS	9.37e-74
7	412	54.4	585	4	008380	MAC-2 BINDING PROTEIN	8.74e-73
8	402	53.1	578	11	P70117	PANCREAS CANCER-ASSOCI	2.30e-70
9	402	53.1	1785	4	09Y211	DMBT1 PROTEIN.	2.30e-70
10	402	53.0	1785	4	09Y4V9	DMBT1/6K.1 PROTEIN PR	2.30e-70
11	401	53.0	504	6	028908	PEPTIDYLPROLYL ISOMERA	4.01e-70
12	397	52.4	918	13	092098	PEMA-SRCR PROTEIN PREC	3.71e-69
13	397	52.4	1594	6	095218	HENSTIN.	3.71e-69
14	395	52.2	574	11	070513	MAMA.	1.13e-68
15	393	51.9	356	6	097682	UNKNOWN MRNA, PARTIAL	3.42e-68
16	385	50.9	2083	11	060997	CRP-DUCIN PRECURSOR	5.04e-66
17	384	50.7	577	11	007797	PEPTIDYLPROLYL ISOMERA	8.76e-66
18	383	50.6	369	6	029110	SCAVENGER-RECEPTOR PRO	8.76e-66
19	383	50.6	895	13	096M1	LYSL OXIDASE RELATED	8.76e-66
20	383	50.6	1036	5	097378	SCAVENGER RECEPTOR CYS	8.76e-66

ALIGNMENTS

RESULT	ID	Query Match	Score	DB	Length	Matches	Mismatches	Indels	Gaps
1	096943	59.4%	450	5	2043	57	20	0	0
2	097375	56.4%	427	5	2153	57	20	0	0
3	007898	55.0%	416	4	1116	57	20	0	0
4	007901	55.0%	416	4	1149	57	20	0	0
5	007899	55.0%	416	4	1151	57	20	0	0
6	007900	55.0%	416	4	1156	57	20	0	0
7	008380	54.4%	412	4	585	57	20	0	0
8	P70117	53.1%	402	11	578	57	20	0	0
9	09Y211	53.1%	402	4	1785	57	20	0	0
10	09Y4V9	53.0%	402	4	1785	57	20	0	0
11	028908	53.0%	401	6	504	57	20	0	0
12	092098	52.4%	397	13	918	57	20	0	0
13	095218	52.4%	397	6	1594	57	20	0	0
14	070513	52.2%	395	11	574	57	20	0	0
15	097682	51.9%	393	6	356	57	20	0	0
16	060997	50.9%	385	11	2083	57	20	0	0
17	007797	50.7%	384	11	577	57	20	0	0
18	029110	50.6%	383	6	369	57	20	0	0
19	096M1	50.6%	383	13	895	57	20	0	0
20	097378	50.6%	383	5	1036	57	20	0	0

```

DE SCVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.
GN SCRI2.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RA PANCER Z., RAST J., DAVIDSON E.H.;
RT "Evolution of the Immune System: Transcription Factors and Downstream
RT Genes of the Mammalian Immune System Expressed in Sea Urchin
RT Coelomocytes."
RL Immunogenetics 0:0-0(1999).
DR EMBL; AF064259; AAD08654.1; -.
DR PROSITE; PS00420; SPERACT_RECEPTOR. 15.
KW Signal; Receptor.
FT SIGNAL 1 16
CHAIN 17 2153
SEQUENCE 2153 AA; 226557 MW; CA3BA8A9 CRC32;

Query Match 56.4%; Score 427; DB 5; Length 2153;
Best Local Similarity 55.4%; Pred. No. 2,00e-76;
Matches 56; Conservative 23; Mismatches 22; Indels 0; Gaps 0;

DB 462 VRLVGLNREGREVEIFLNNQGTVCDDMGTPDANVYCRQLGYPGSGARSSAYFGRGS 521
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQWDADAELVCRQLGSLAKAMHQAIFEGGS 286
DB 522 VILLIDNVCCSGNERSLSCNNGICVHNGCHQEDASVYCT 562
QY 287 GPVMDDEVRCITGNELSTEQCPKSSWGEHNGHKEDAGVSC 327

RESULT 3
ID 007898 PRELIMINARY; PRT; 1116 AA.
AC 007898.
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN PRECURSOR.
GN CD163.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA IAN S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily."
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL; Y18388; CAB45233.1; -.
DR EMBL; Y18389; CAB45233.1; JOINED.
DR EMBL; Y18390; CAB45233.1; JOINED.
DR EMBL; Y18391; CAB45233.1; JOINED.
DR EMBL; Y18392; CAB45233.1; JOINED.
DR EMBL; Y18393; CAB45233.1; JOINED.
DR EMBL; Y18394; CAB45233.1; JOINED.
DR EMBL; Y18395; CAB45233.1; JOINED.
DR EMBL; Y18396; CAB45233.1; JOINED.
DR EMBL; Y18397; CAB45233.1; JOINED.
DR EMBL; Y18398; CAB45233.1; JOINED.
DR EMBL; Y18399; CAB45233.1; JOINED.
DR EMBL; Y18400; CAB45233.1; JOINED.
DR EMBL; Y18401; CAB45233.1; JOINED.

```

```

DR EMBL; Y18402; CAB45233.1; JOINED.
DR EMBL; Y18403; CAB45233.1; JOINED.
DR PFAM; PF00530; SCRC; 9.
DR PRINTS; PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
CHAIN 41 1116
SEQUENCE 1116 AA; 120979 MW; F193FBBA CRC32;

Query Match 55.0%; Score 416; DB 4; Length 1116;
Best Local Similarity 55.4%; Pred. No. 9,37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

DB 924 IRLQEGFTSCSGRVELYHGGSGWTCVDDSDVDDAQQVCOQLGCPALAKFAKFAEFGGT 983
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQWDADAELVCRQLGSLAKAMHQAIFEGGS 286
DB 984 GPVMDDEVRCITGNELSTEQCPKSSWGEHNGHKEDAGVSC 327
QY 287 GPVMDDEVRCITGNELSTEQCPKSSWGEHNGHKEDAGVSC 327

RESULT 4
ID 007901 PRELIMINARY; PRT; 1149 AA.
AC 007901.
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, EXTRACELLULAR VARIANT PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA IAN S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily."
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL; Z22971; CAB80544.1; -.
DR PFAM; PF00530; SCRC; 9.
DR PRINTS; PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
CHAIN 41 1149
SEQUENCE 1149 AA; 124328 MW; 4901C708 CRC32;

Query Match 55.0%; Score 416; DB 4; Length 1149;
Best Local Similarity 55.4%; Pred. No. 9,37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

DB 957 IRLQEGFTSCSGRVELYHGGSGWTCVDDSDVDDAQQVCOQLGCPALAKFAKFAEFGGT 1016
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDQWDADAELVCRQLGSLAKAMHQAIFEGGS 286
DB 1017 GPVMDDEVRCITGNELSTEQCPKSSWGEHNGHKEDAGVSC 327
QY 287 GPVMDDEVRCITGNELSTEQCPKSSWGEHNGHKEDAGVSC 327

RESULT 5
ID 007899 PRELIMINARY; PRT; 1151 AA.
AC 007899.
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE: 93380506.
RA LAM S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: 222968; CA80542.1; -.
DR PFAM: PF00530; SRCR: 9; -.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1151
SQ SEQUENCE 1151 AA; 124820 MW; A72EDD2F CRC32;
POTENTIAL.
MI30 ANTIGEN, CYTOPLASMIC VARIANT 1.
Query Match 55.0%; Score 416; DB 4; Length 1151;
Best Local Similarity 55.4%; Pred. No. 9,37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;
DB 924 IRLQEGPTSCSGRVELYHGGSGMTVCDDSDLDADQVYCCQLGCGPALKAFKREAGGCT 983
QY 227 IRLAGSSVHEGRVELYHAGQMTVCDDDDADAEVICHQLSLGIAKAMHQAIFYEGGS 286
DB 984 GPIMLNEVKCKGNSSLMDCPARRMGSHGKEDAAVNCCT 1024
QY 287 GPVMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 327
RESULT 6 PRELIMINARY; PRT; 1156 AA.
ID 007900;
AC 007900;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE M130 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 93380506.
RA LAM S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the
RT scavenger receptor superfamily";
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: 222970; CA80543.1; -.
DR PFAM: PF00530; SRCR: 9; -.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1156
SQ SEQUENCE 1156 AA; 125352 MW; 287A07A0 CRC32;
POTENTIAL.
MI30 ANTIGEN, CYTOPLASMIC VARIANT 2.
Query Match 55.0%; Score 416; DB 4; Length 1156;
Best Local Similarity 55.4%; Pred. No. 9,37e-74;
Matches 56; Conservative 19; Mismatches 26; Indels 0; Gaps 0;
DB 924 IRLQEGPTSCSGRVELYHGGSGMTVCDDSDLDADQVYCCQLGCGPALKAFKREAGGCT 983
QY 227 IRLAGSSVHEGRVELYHAGQMTVCDDDDADAEVICHQLSLGIAKAMHQAIFYEGGS 286
DB 984 GPIMLNEVKCKGNSSLMDCPARRMGSHGKEDAAVNCCT 1024
QY 287 GPVMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 327
RESULT 7 PRELIMINARY; PRT; 585 AA.
ID 008380;
AC 008380;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE MAC-2 BINDING PROTEIN PRECURSOR.

OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 93300818.
RA KOTHS K., TAYLOR E., HALENBECK R., CASIPIT C., WANG A.;
RT "Cloning and characterization of a human Mac-2-binding protein, a new
RT member of the superfamily defined by the macrophage scavenger receptor
RT cysteine-rich domain";
RL J. Biol. Chem. 268:14245-14249(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE: 94308070.
RA ULLRICH A., SURES I., D'EGIDO M., JALIAL B., POWELL T.J., HERBST R.,
RA DREPS A., AZAM M., RUBINSTEIN M., NATOLI C.;
RT "The secreted tumor-associated antigen 90K is a potent immune
RT stimulator";
RL J. Biol. Chem. 269:18401-18407(1994).
DR EMBL: 143210; AAA36193.1; -.
DR EMBL: X79089; CAA55699.1; -.
DR PFAM: PF00530; SRCR: 1; -.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 585
SQ SEQUENCE 585 AA; 65330 MW; AAE9E32 CRC32;
POTENTIAL.
MAC-2 BINDING PROTEIN.
Query Match 54.4%; Score 412; DB 4; Length 585;
Best Local Similarity 53.5%; Pred. No. 8,74e-73;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;
DB 24 MRLADGATNCGRVELYFFYGGMTVCDDMDLTDASVCRALGFENATQALGRAFGGS 83
QY 227 IRLAGSSVHEGRVELYHAGQMTVCDDDDADAEVICHQLSLGIAKAMHQAIFYEGGS 286
DB 84 GPIMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 124
QY 287 GPVMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 327
RESULT 8 PRELIMINARY; PRT; 578 AA.
ID P70117;
AC P70117;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PANCREAS CANCER-ASSOCIATED PROTEIN 4.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
RP SEQUENCE FROM N.A.
RA SCHAFER C., POOR P.M., MACDONALD R.G., CHANEY W.G.;
RT Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U73375; AAB18745.1; -.
DR PROSITE: PS00420; SPERACTRCPTR.
DR PFAM: PF00530; SRCR: 1; -.
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 578
SQ SEQUENCE 578 AA; 64398 MW; 002E8BD2 CRC32;
POTENTIAL.
Query Match 53.1%; Score 402; DB 11; Length 578;
Best Local Similarity 52.5%; Pred. No. 2,30e-70;
Matches 53; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
DB 24 MRLVNGASANEGRVELYFGQMTVCDDMDLTDANVYCRALGFENATQALGRAFGGR 83
QY 227 IRLAGSSVHEGRVELYHAGQMTVCDDDDADAEVICHQLSLGIAKAMHQAIFYEGGS 286
DB 84 GPIMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 124
QY 287 GPVMLDEVCTGNELSLIEQCPKSSWGEHNGHKEDAGVST 327

ID	RESULT	9		PRT:	1785 AA.
AC	O9y211		PRELIMINARY;		
DT	01-NOV-1999	(TrEMBLrel. 12,			
DT	01-NOV-1999	(TrEMBLrel. 12,			
DT	01-NOV-1999	(TrEMBLrel. 12,			
DE	DMBT1 PROTEIN.				
GN	DMBT1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
CC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
RN	[1]				
RA	HORII A.;				
RP	SEQUENCE FROM N.A.				
RT	"DMBT1."				
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
	EMBL: AB020851; BAA78577.1; .				
	EMBL: AB020812; BAA78577.1; JOINED.				
	EMBL: AB020813; BAA78577.1; JOINED.				
	EMBL: AB020814; BAA78577.1; JOINED.				
	EMBL: AB020816; BAA78577.1; JOINED.				
DR	EMBL: AB020817; BAA78577.1; JOINED.				
DR	EMBL: AB020818; BAA78577.1; JOINED.				
DR	EMBL: AB020819; BAA78577.1; JOINED.				
DR	EMBL: AB020820; BAA78577.1; JOINED.				
DR	EMBL: AB020821; BAA78577.1; JOINED.				
DR	EMBL: AB020822; BAA78577.1; JOINED.				
DR	EMBL: AB020823; BAA78577.1; JOINED.				
DR	EMBL: AB020824; BAA78577.1; JOINED.				
DR	EMBL: AB020825; BAA78577.1; JOINED.				
DR	EMBL: AB020826; BAA78577.1; JOINED.				
DR	EMBL: AB020827; BAA78577.1; JOINED.				
DR	EMBL: AB020828; BAA78577.1; JOINED.				
DR	EMBL: AB020829; BAA78577.1; JOINED.				
DR	EMBL: AB020830; BAA78577.1; JOINED.				
DR	EMBL: AB020831; BAA78577.1; JOINED.				
DR	EMBL: AB020832; BAA78577.1; JOINED.				
DR	EMBL: AB020833; BAA78577.1; JOINED.				
DR	EMBL: AB020834; BAA78577.1; JOINED.				
DR	EMBL: AB020835; BAA78577.1; JOINED.				
DR	EMBL: AB020836; BAA78577.1; JOINED.				
DR	EMBL: AB020837; BAA78577.1; JOINED.				
DR	EMBL: AB020838; BAA78577.1; JOINED.				
DR	EMBL: AB020839; BAA78577.1; JOINED.				
DR	EMBL: AB020840; BAA78577.1; JOINED.				
DR	EMBL: AB020841; BAA78577.1; JOINED.				
DR	EMBL: AB020842; BAA78577.1; JOINED.				
DR	EMBL: AB020843; BAA78577.1; JOINED.				
DR	EMBL: AB020844; BAA78577.1; JOINED.				
DR	EMBL: AB020845; BAA78577.1; JOINED.				
DR	EMBL: AB020846; BAA78577.1; JOINED.				
DR	EMBL: AB020847; BAA78577.1; JOINED.				
DR	EMBL: AB020848; BAA78577.1; JOINED.				
DR	EMBL: AB020849; BAA78577.1; JOINED.				
DR	EMBL: AB020850; BAA78577.1; JOINED.				
SQ	SEQUENCE 1785 AA; 193991 MW; 479F75D8 CRC32;				

Query Match 53.1%; Score 402; DB 4; Length 1785;
 Best Local Similarity 51.5%; Pred. No. 2,30e+70;
 Matches 52; Conservative 19; Mismatches 30; Indels 0; Gaps

Db	1258 UNLSSYGLCAGREIYHGSTGWTGVDDSDSTIOEAIVGRQLGCGRAYSALGNATFGSGS	1317
Gy	227 IRLNGSVSHVEGRRIELTHAQOWKTVCDQDDQADADAIVICRQLGSLGIAMQHQAAYFEGS	286
Db	1318 GPITLDVPCSGSTESTLMQCRRNGWFSHCNHRFDAGVICS	1358
Gy	287 GPVMLDEVRCIGNELSTEGCPKSKSWBHNGHGHRDAGVCT	327

RESULT 10

```

ID AC 09Y4V9 PRELIMINARY: PRT, 1785 AA.
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE DBMT1/6KB.1 PROTEIN PRECURSOR.
GN DBMT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG.
RX MEDLINE: 97434209.
RA MOLLENAER J., WIEMANN S., SCHEUREN W., KORN B., HAYASHI Y.,
RA WILGENBUS K.K., VON DEMLING A., POUSTKA A.;
RT "DBMT1, a new member of the SRCR superfamily, on chromosome 10q25.3-
RT 26.1 is deleted in malignant brain tumours.";
RL Nat. Genet. 17:32-39(1997).
DR EMBL: A000342; CAA04019.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN
FT CHAIN 26 1785 DBMT1/6KB.1 PROTEIN.
SQ SEQUENCE 1785 AA; 193941 MW; 9259828E CRC32;

Query Match 53.1%; Score 402; DB 4; Length 1785;
Best Local Similarity 51.5%; Pred. No. 2.30e-70;
Matches 52; Conservative 19; Mismatches 30; Indels 0; Gaps 0

Db 1258 UNLSSVGLCAGVEYEHGCTMGTCVCDSDSWIQAFAEYVCROLGCGRAVSAALGNAYFGSGS 1317
Oy 227 IRLAGSSVHGVEGVEYHYHAGQGVCDQWDADAIEYICROLGSIARAKHQAAYFGS 286
Db 1318 GPILDDVEGSGTESTLMQCRNRGWFNSHNCNRREDAGVICS 1358
Oy 287 GPVMDDEVRCGTGNELSIEGCPRSSNGEHCNRKRDAGVST 327

RESULT 11
ID 028908 PRELIMINARY: PRT, 504 AA.
AC 028908;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE MOCIN (FRAGMENT).
DE Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95374471.
RA NINES D.P., KEATES A.C., AFDAHL N.H., OFFNER G.D.;
RT "Bovine gall-bladder mucin contains two distinct tandem repeating
RT sequences: evidence for scavenger receptor cysteine-rich repeats.";
RL Biochem. J. 310:41-48(1995).
DR EMBL: S78869; AAB35070.1; -.
DR PRAM: PF00431; CUB: 1.
DR PRAM: PF00530; SRCR: 1.
FT NON_TER 1 504
FT NON_TER 1 504
SQ SEQUENCE 504 AA; 54864 MW; EF298513 CRC32;

Query Match 53.0%; Score 401; DB 6; Length 504;
Best Local Similarity 60.4%; Pred. No. 4.01e-70;
Matches 55; Conservative 13; Mismatches 21; Indels 2; Gaps 2.

Db 328 GREYIEHGQOMCTVCDMDWDQDAQVVCRLG-CGYAVSAPGNAYFGSGSPITLDDVYC 386
Oy 238 GREYELIHAQOMGTVCDDQWDADAIEYICROLGLSLAKAH-QAYTGESSGPVMDDEVRC 296
Db 387 SGAESNLMQCRNRGWFNSHNCNRREDAGVICS 417

```


Thu Mar 16 07:57:57 2000

US-09-147-947-6-04.rsp

Page 6

Qy 227 IRLAGSSVHGGRVELYIAGMGIVCDQDDMDADAEVICRGLSGIAKMAHAYFEBS 286
Dd 200 GPITLDVECSGSESTLMQCNRNGFSHNCRHNEDAGVICS 240
||: ||: | |: | : : | : : | ||: ||: ||| |:
Qy 287 GPMVLDEVRCTGNELSIQCPSKSSGGEINCHCKEKEDAGVSCN 327

```
Search completed: Mon Mar 13 10:28:00 2000
Job time : 18 secs.
```



```
QY 176 LPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 216

RESULT 2
ENTRY 114893 #type complete
TITLE scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)
ORGANISM #formal_name Strongylocentrotus purpuratus #common_name purple urchin
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS 114893
REFERENCE 218253
AUTHORS Pancer, Z.; Rast, J.P.; Davidson, E.H.
JOURNAL Immunogenetics (1999) 49:773-786
#title Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.

#accession 114893
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-2153 #label PAN
#cross-references EMBL:AF064259; NID:g4165052; EID:g4165053; PID:AA00654.1

GENETICS
#gene SRCR12
SUMMARY #length 2153 #molecular-weight 226556 #checksum 3510

Query Match
Best Local Similarity 38.6%; Pred. No. 3,72e-39; Length 2153;
Matches 39; Conservative 22; Mismatches 40; Indels 0; Gaps 0;

Db 462 VRLVGLNLRGRVIFLNNMGVTCDDMGTPDANVVCROLGPGSGARSAYFRGS 521
117 VRLRGKNEFEETVYASGVWGTVCSSHMWDSASVICHQLQLGKIAKQPFPSGLG 176
QY 522 VPILDNVCSSGERSLECSNNGICVHNCGHQEDASVYCT 562
177 IPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 217

RESULT 3
ENTRY 146862 #type complete
TITLE macrophage scavenger receptor type I - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 20-Sep-1999

ACCESSIONS 146862
REFERENCE 146862
AUTHORS Bickel, P.E.; Freeman, M.W.
JOURNAL J. Clin. Invest. (1992) 90:1450-1457
#title Rabbit aortic smooth muscle cells express inducible macrophage scavenger receptor messenger RNA that is absent from endothelial cells.

#cross-references MUID:93016877
#accession 146862
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-454 #label BIC
#cross-references GB:11193; NID:g165508; PID:AAA31402.1; PID:g165509
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor cysteine-rich domain homology

FEATURE 350-453
SUMMARY #domain scavenger receptor cysteine-rich domain homology
#length 454 #molecular-weight 49745 #checksum 2903

Query Match
Best Local Similarity 40.8%; Pred. No. 7.99e-37; Length 454;
Matches 42; Conservative 17; Mismatches 40; Indels 4; Gaps 4;

Db 353 VRLVGLNLRGRVIFLNNMGVTCDDMGTPDANVVCROLGPGSGARSAYFRGS 521
177 IPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 217

QY 117 VRLRGKNEFEETVYASGVWGTVCSSHMWDSASVICHQLQLGKIAKQPFPSGLG 175
176 LPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 217

RESULT 4
ENTRY 536077 #type complete
TITLE M130 antigen - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Aug-1996

ACCESSIONS 536077
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.

#accession 138003
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-1116 #label RES
#cross-references EMBL:Z22968; NID:g312141; PID:g312142
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 43-147

Db 412 TGPI-WLNEVPCLGMSIECKIRQWGVYRCSHGEDAGVTC 453
176 LPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 217
QY 176 LPIYMSNVRGRGDENILLCERKIDIMOGVCPOKMAAAVTC 217

RESULT 5
ENTRY 138006 #type complete
TITLE M130 antigen (extracellular variant) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 02-Aug-1996

ACCESSIONS 138006
REFERENCE 138003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
JOURNAL Eur. J. Immunol. (1993) 23:2320-2325
#title A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
```


QY 117 VRLRGKNEFEGETVEYVAGWGVCSHMDSDASVICHOLQLOGKGLA-KOTPFSGLG 175

Db 983 TGPI-WLNEVCKCKGNESLMDPCPARMGHSECGHEDAAVNCT 1024

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWGGVCPQKMAAAVTC 217

RESULT 8

ENTRY A44407 #type complete
TITLE macrophage scavenger receptor, RSRI (collagen-like domain) - rabbit

ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

ACCESSIONS A44407
REFERENCE A44407
#authors Doi, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.; Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.; Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
#journal J. Biol. Chem. (1993) 268:2126-2133
#title Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors.

#cross-references MUID:93131972
#accession A44407
#status preliminary: not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-454 #label DOI
##experimental_source lung
##note sequence extracted from NCBI backbone (NCBIP:123206)
CLASSIFICATION #superfamily unassigned collagens: scavenger receptor
cysteine-rich domain homology

FEATURE 350-453
#domain scavenger receptor cysteine-rich domain homology
SUMMARY #label SRC
#length 454 #molecular-weight 49735 #checksum 2435

Query Match 32.5%; Score 248; DB 2; Length 454;
Best Local Similarity 39.8%; Pred. No. 5,71e-35;
Matches 41; Conservative 17; Mismatches 41; Indels 4; Gaps 4;

Db 353 VRLVGGSGHGRVREILHNGMGVCDHWEIRAGOVCRSLGYGVSVAKKAF-GQG 411

QY 117 VRLRGKNEFEGETVEYVAGWGVCSHMDSDASVICHOLQLOGKGLA-KOTPFSGLG 175

Db 412 TGPI-WLNEVPCLEMESSIECKIRMGVRCVSHGEDAGVCT 453

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWGGVCPQKMAAAVTC 217

ENTRY 9
TITLE B44407 #type complete
ALTERNATE_NAMES macrophage scavenger receptor, MSRI - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
ACCESSIONS B44407; A38260; I56334
REFERENCE A44407
#authors Doi, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.; Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.; Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
#journal J. Biol. Chem. (1993) 268:2126-2133
#title Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors.

#cross-references MUID:93131972
#accession B44407
#status preliminary: not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-458 #label DOI
##experimental_source macrophage-like cell line P388D

REFERENCE #note
A38260 sequence extracted from NCBI backbone (NCBIP:123207)

#authors Freeman, M.; Ashkenas, J.; Rees, D.J.G.; Kingsley, D.M.; Copeland, N.G.; Jenkins, N.A.; Krieger, M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8810-8814
#title An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors.

#cross-references MUID:91062370
#accession A38260
##status preliminary
##molecule_type mRNA
#residues 349-458 #label FRE
#cross-references GB:M59445; GB:M36817; NID:g192737; PIDN:AAA37464.1; PID:g192738

REFERENCE I56334
#authors Ashkenas, J.; Penman, M.; Vassile, E.; Acton, S.; Freeman, M.W.; Krieger, M.
#journal J. Lipid Res. (1993) 34:983-1000
#title Structures and high and low affinity ligand binding properties of murine type I and type II macrophage scavenger receptors.

#cross-references MUID:93359822
#accession I56334
##status preliminary: translated from GB/EMBL/DBJ
##molecule_type mRNA
#residues 5-458 #label RES
#cross-references GB:I04274; NID:9293745; PIDN:AAA39747.1; PID:g293746
CLASSIFICATION #superfamily unassigned collagens: scavenger receptor
cysteine-rich domain homology

FEATURE 354-457
#domain scavenger receptor cysteine-rich domain homology
SUMMARY #label SRC
#length 458 #molecular-weight 50130 #checksum 1435

Query Match 31.9%; Score 244; DB 2; Length 458;
Best Local Similarity 36.9%; Pred. No. 4,79e-34;
Matches 38; Conservative 19; Mismatches 42; Indels 4; Gaps 4;

Db 357 VRLVGSAGHGRVREIFRFGOGKFTICDDRDIRAGOVCRSLGYEVLAVKRAHF-GQG 415

QY 117 VRLRGKNEFEGETVEYVAGWGVCSHMDSDASVICHOLQLOGKGLA-KOTPFSGLG 175

Db 416 TGPI-WLNEVPCFESESIENCKINQWGVLSCHSEDAVCT 457

QY 176 LIPIYW-SNVRGRGDEENILLCEKDIWGGVCPQKMAAAVTC 217

RESULT 10

ENTRY A55840 #type complete
TITLE macrophage bacteria-binding receptor MARCO - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 16-Jul-1999

ACCESSIONS A55840
REFERENCE A55840
#authors Elomaa, O.; Kangas, M.; Sahlberg, C.; Tuukkanen, J.; Sormunen, R.; Liakka, A.; Thesleff, I.; Kraal, G.; Tryggvason, K.
#journal Cell (1995) 80:603-609
#title Cloning of a novel bacteria-binding receptor structurally related to scavenger receptors and expressed in a subset of macrophages.

#cross-references MUID:95171455
#accession A55840
##status preliminary
##molecule_type mRNA
#residues 1-518 #label ELO
#cross-references GB:U18424; NID:9682722; PID:9682723
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS transmembrane protein
FEATURE 418-518
#domain scavenger receptor cysteine-rich domain homology

117 VRLRGKNEFTEVYASGVWGTVCSSHWDDSDASVICHQLQLG-GKGIKQTPPSGLG 175
Db 1214 NGSI-WLDEVOGCGRESSLMDCAEPWPGSDCKHEPDAGVRC 1255
176 LPIYV-SNVRGRGDENILLCEKDIWGGVCPQKMAAVTCS 217

RESULT 13
ENTRY S56744 #type fragment
TITLE mucin (clone PGM7-1) - bovine (fragment)
ORGANISM #formal name Bos primigenius taurus #common name cattle
DATE 27-Oct-1995 #sequence-revision 30-Jan-1998 #text-change 07-May-1999

ACCESSIONS S56744
REFERENCE S56744
#authors Nunes, D.P.; Keates, A.C.; Afhal, N.H.; Offner, G.D.
#journal Biochem. J. (1995) 310:41-48
#title Bovine gall-bladder mucin contains two distinct tandem repeating sequences: evidence for scavenger receptor cysteine-rich repeats.
#cross-references MUID:95374471
#accession S56744
#molecule-type mRNA
#residues 1-600 #label NUN

CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE 1-102
126-229 #domain scavenger receptor cysteine-rich domain homology (fragment) #label SRC1\
253-356 #domain scavenger receptor cysteine-rich domain homology #label SRC2\
380-483 #domain scavenger receptor cysteine-rich domain homology #label SRC3\
#domain scavenger receptor cysteine-rich domain homology #label SRC4

SUMMARY #length 600 #checksum 7824

Query Match 30.6%; Score 234; DB 2; Length 600;
Best Local Similarity 38.2%; Pred. No. 9,49e-32;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

Db 383 LRLVNGSDRCGRFVILVGGSMGTCDSDMDTNDANVVCRGSGSISAPGDARF-GGG 441
QY 117 VRLRGKNEFTEVYASGVWGTVCSSHWDDSDASVICHQLQLG-GKGIKQTPPSGLG 175
Db 442 SGPIILDVCGSGVETVYMSCSHSPWNSHCHSKDASYICS 483
QY 176 LPIYVSNVRGRGDENILLCEKDIWGGVCPQKMAAVTCS 217

14
S08276 #type complete
LDL receptor I, macrophage - bovine
#formal name Bos primigenius taurus #common name cattle
DATE 30-Jun-1992 #sequence-revision 30-Jun-1992 #text-change 20-Sep-1999

ACCESSIONS S08276
REFERENCE S08276
#authors Kodama, T.; Freeman, M.; Rohrer, L.; Zabrecky, J.; Matsuda, P.; Krieger, M.
#journal Nature (1990) 343:531-535
#title Type I macrophage scavenger receptor contains alpha-helical and collagen-like coiled coils.
#cross-references MUID:90136965
#accession S08276
#status not compared with conceptual translation
#molecule-type mRNA
#residues 1-453 #label KOD
#cross-references GB:X51689; GB:X54182; NID:9734; PIDN:CAA35987.1; PID:9735

CLASSIFICATION #superfamily unassigned collagens; scavenger receptor

KEYWORDS cysteine-rich domain homology
alternative splicing; coiled coil; glycoprotein; transmembrane protein
FEATURE 51-76
349-452 #domain transmembrane #status predicted #label TMN\
#domain scavenger receptor cysteine-rich domain homology #label SRC\
82,101,142,183,220, #binding-site carbohydrate (Asn) (covalent) #status predicted
248,266

SUMMARY #length 453 #molecular-weight 50056 #checksum 4303

Query Match 30.5%; Score 233; DB 2; Length 453;
Best Local Similarity 37.9%; Pred. No. 1.61e-31;
Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 352 VRLVGGSGHEBRVLEFPHGCGMTCDRWELRGVLVCRSLGKYGVSVKRAVFG-GKG 410
QY 117 VRLRGKNEFTEVYASGVWGTVCSSHWDDSDASVICHQLQLG-GKGIKQTPPSGLG 175
Db 411 TGPV-WLNEVFCGKRESSIECRIRQMGVRACSHDEDAVTC 452
QY 176 LPIYV-SNVRGRGDENILLCEKDIWGGVCPQKMAAVTCS 217

RESULT 15
ENTRY JC4361 #type complete
TITLE scavenger receptor Cys-rich epidermal growth factor precursor - sea lamprey
ORGANISM #formal name Petromyzon marinus #common name sea lamprey
DATE 08-Jan-1996 #sequence-revision 08-Feb-1996 #text-change 16-Jul-1999

ACCESSIONS JC4361
REFERENCE JC4361
#authors Mayer, W.E.; Tichy, H.
#journal Gene (1995) 164:267-271
#title A cDNA clone from the sea lamprey Petromyzon marinus coding for a scavenger receptor Cys-rich (SRCR) domain protein.
#cross-references MUID:96069593
#accession JC4361

CLASSIFICATION #molecule-type mRNA
#residues 1-918 #label MAY
#cross-references GB:U20652; NID:9790233; PID:9790234

COMMENT This protein is rich in cysteine and plays a role in intercellular contacts and cell activation or differentiation in the cyclosome immune system.

CLASSIFICATION #superfamily EGF homology; scavenger receptor cysteine-rich domain homology
glycoprotein; growth factor; receptor; transmembrane protein

KEYWORDS
FEATURE 1-24
25-918
55-157
179-212
234-267
295-328
356-389
410-443
463-565
876-895
896-918
40,545,575,585,
814

SUMMARY #length 918 #molecular-weight 101417 #checksum 2936

Query Match 30.1%; Score 230; DB 2; Length 918;
Best Local Similarity 39.8%; Pred. No. 7.79e-31;
Matches 41; Conservative 17; Mismatches 40; Indels 5; Gaps 5;

Thu Mar 16 07:57:49 2000

US-09-147-947-6-03.rpr

Page 7

Db 466 VRLVGGGV-CGGRREVVYVYAGSMGVICDDSDMDROAEVYCGVYVSVASPSNAHF-GVG 523
117 VRLGGRKNEFEFGYEVYASGVYGVICSSHHDDSDASVICHQLQLQG-KGIARLQPTPSGLG 175

Db 524 SGQI-WLDDVYVCGEESLACNINSGRHHCGNEDASVYCS 565

Qy 176 LPIIYV-SVAVRCRDEENILCEMDIQQGVYCPKMAAAVYCS 217

Search completed: Mon Mar 13 10:25:18 2000
Job time : 14 secs.

THIS PAGE BLANK (08PT0)

(J.M.)

Distribution rights by Oxford Molecular Ltd

1 VRLRGKNEFEGTVEVYASG.....KDIWQGGVC P Q K M A A V T C S 101

Gap 11

131253 seqs, 12956647 residues

Listing first 45 summaries

1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

mean 26.677; variance 101.261; scale 0.263

and is derived by analysis of the total score distribution.

SUMMARIES

	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	239	31.3	489	2	US-08-794-	Sequence 7, Applicatio	6.02e-15
2	239	31.3	518	1	US-08-892-	Sequence 2, Applicatio	6.02e-15
3	237	31.0	451	1	PCT-US96-0	Sequence 2, Applicatio	9.70e-15
4	237	31.0	451	1	US-08-453-	Sequence 2, Applicatio	9.70e-15
5	237	31.0	451	2	US-08-948-	Sequence 2, Applicatio	9.70e-15
6	237	31.0	451	2	US-08-973-	Sequence 2, Applicatio	9.70e-15
7	237	31.0	451	1	US-08-154-	Sequence 2, Applicatio	9.70e-15
8	224	29.3	453	4	5510466-4	Patent No. 5510466.	2.13e-13
9	219	28.7	585	2	US-08-316-	Sequence 10, Applicati	6.96e-13
10	219	28.7	585	1	US-08-477-	Sequence 10, Applicati	6.96e-13
11	219	28.7	585	1	US-08-473-	Sequence 10, Applicati	6.96e-13
12	217	28.4	495	2	US-08-794-	Sequence 2, Applicatio	1.12e-12
13	217	28.4	520	2	US-08-794-	Sequence 6, Applicatio	1.12e-12
14	212	27.7	1390	1	US-08-770-	Sequence 2, Applicatio	3.63e-12
15	113	14.8	798	1	US-08-800-	Sequence 2, Applicatio	1.93e-02
16	113	14.8	798	3	PCT-US94-0	Sequence 2, Applicatio	1.93e-02
17	88	11.5	1130	4	5444156-2	Patent No. 5444156.	3.31e+00
18	88	11.5	1130	4	US-08-125-	Sequence 2, Applicatio	3.31e+00
19	88	11.5	1130	2	US-08-125-	Sequence 2, Applicatio	3.31e+00
20	88	11.5	1311	2	US-08-160-	Sequence 4, Applicatio	3.31e+00
21	88	11.5	1311	2	US-08-160-	Sequence 4, Applicatio	3.31e+00
22	83	10.9	268	2	US-08-453-	Sequence 65, Applicati	8.83e+00
23	82	10.7	1719	2	US-08-459-	Sequence 4, Applicatio	1.08e+01

45	70	9.2	550	3	PCT-US93.0	Sequence 1, Applicatio	1.04e+020
44	70	9.2	456	2	US-08-819	Sequence 1, Applicatio	1.04e+020
43	70	9.2	263	2	US-08-353	Sequence 67, Applicati	1.04e+020
42	70	9.2	263	2	US-08-333	Sequence 67, Applicati	1.04e+020
41	71	9.3	338	2	US-08-333	Sequence 66, Applicati	1.04e+020
40	71	9.3	294	1	US-09-047	Sequence 2, Applicatio	8.62e+010
39	71	9.3	261	1	US-08-654	Sequence 2, Applicatio	8.62e+010
38	72	9.4	74	1	US-08-420	Sequence 2, Applicatio	7.17e+010
37	72	9.4	74	1	US-08-563	Sequence 2, Applicatio	7.17e+010
36	73	9.6	1824	2	US-08-660	Sequence 3, Applicatio	5.96e+010
35	73	9.6	429	3	PCT-US95.0	Sequence 109, Applicat	5.96e+010
34	73	9.6	429	2	US-08-310	Sequence 109, Applicat	5.96e+010
33	73	9.6	422	2	US-08-712	Sequence 2, Applicatio	5.96e+010
32	75	9.8	3729	2	US-08-894	Sequence 4, Applicatio	4.10e+010
31	75	9.8	327	2	US-08-896	Sequence 38, Applicati	4.10e+010
30	75	9.8	223	2	US-08-896	Sequence 4, Applicatio	4.10e+010
29	77	10.1	1297	2	US-08-290	Sequence 4, Applicatio	2.81e+010
28	77	10.1	423	2	US-08-261	Sequence 11, Applicati	2.81e+010
27	77	10.1	356	2	US-08-631	Sequence 1, Applicatio	2.81e+010
26	79	10.3	1706	2	US-08-339	Sequence 2, Applicatio	1.92e+010
25	79	10.3	1706	2	US-08-339	Sequence 2, Applicatio	1.92e+010
24	82	10.7	1719	2	US-08-439	Sequence 4, Applicatio	1.08e+010

ALIGNMENTS

RESULT 1 STANDARD; PRT: 489 AA.
 ID US-08-794-795-7
 AC xxxxxx
 DT :
 XX :
 DE :
 XX Sequence 7, Application US/08794795
 CC /Sequence 7, Application US/08794795
 CC Patent No. 5916766
 CC GENERAL INFORMATION:
 CC APPLICANT: Eishourlagy, Nabii
 CC APPLICANT: Adamou, John
 CC APPLICANT: Gross, Mitchell
 CC APPLICANT: Iysko, Paul
 CC TITLE OF INVENTION: Human Macro Scavenger Re
 CC TITLE OF INVENTION: epcor
 CC NUMBER OF SEQUENCES: 9
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Smithkline Beecham Corporation
 CC STREET: 709 Swedeland Road
 CC City: King of Prussia
 CC STATE: PA
 CC COUNTRY: USA
 CC ZIP: 19406
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM Compatible
 CC OPERATING SYSTEM: DOS
 CC SOFTWARE: FASTSEQ for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/794,795
 CC FILING DATE: 04-FEB-1997
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: ATG50009P
 CC FILING DATE: 22-MAY-1996
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Han, William T
 CC REGISTRATION NUMBER: 34,344
 CC REFERENCE/DOCKET NUMBER: ATG50009
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 610-270-5219
 CC TELEFAX: 610-270-4026
 CC TELEX:
 CC INFORMATION FOR SEQ ID NO: 7:

[illegible]

```

CC      TPE: amino acid
CC      STRANDEDNESS: Single
CC      TOPOLOGY: Linear
CC      SEQUENCE 518 AA; 52730 MW; 1281000 CN;
SQ
Query Match          31.3%; Score 239; DB 1; Length 518;
Best Local Similarity 39.2%; Pred. No. 6,02e-15;
Matches 40; Conservative 21; Mismatches 34; Indels 7; Gaps 6
Dd      423 VRIMGGINR--GRAEYVYNNEMGITCDDMDNDATVFCRMIGYS-RCRA-LSSYGG-GS 477
Oy      117 VRLRGKNEFEFGTEVYASGVWGTCSHMDSDASVICHOLQIGKGIAKQTFPSGIGL 176
Dd      418 GNI-WLDVNCRGTENSLMDCSKSWGHHNCVHNEDEGVS 518
Oy      177 IPIYW-SNVRGRGDENILLCEKIDWOGVCPQKMAAVTCS 217
RESULT 3
ID      PCT-US96-08081-2 STANDARD: PRT; 451 AA.
Ac      xxxxxx
Df      Sequence 2, Application PC/TUS9608081
XX      Sequence 2, Application PC/TUS9608081
XX      GENERAL INFORMATION:
CC      APPLICANT: SmithKline Beecham Corporation
CC      TITLE OF INVENTION: Attachment Enhanced 293 Cells
CC      NUMBER OF SEQUENCES: 4
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: SmithKline Beecham - Corporate Patents U.S.
CC      STREET: Mailcode - UW2220, 709 Swedeland Road
CC      CITY: King of Prussia
CC      STATE: Pennsylvania
CC      COUNTRY: U.S.A.
CC      ZIP: 19406-5090
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US96/08081
CC      FILING DATE:
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Jervis, Herbert H.
CC      REGISTRATION NUMBER: 31,171
CC      REFERENCE/DOCKET NUMBER: P50338
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (610) 270-5019
CC      TELEFAX: (610) 270-5090
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 451 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
SQ      SEQUENCE 451 AA; 49762 MW; 1024670 CN;
Query Match          31.0%; Score 237; DB 3; Length 451;
Best Local Similarity 37.9%; Pred. No. 9.70e-15;
Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;
Dd      350 VRLVGSGSPHGHRREILHSGQWGTICDDRMVEYRGQVVCRSILGPGVAAVKAHF-GQG 408
Oy      117 VRLRGKNEFEFGTEVYASGVWGTCSHMDSDASVICHOLQIGKGIAKQTFPSGIGL 175
Dd      409 TGPI-WLNEVFCFGRRESSIECKTRQMGTRACSHSEDAGYTCT 450
Oy      117 VRLRGKNEFEFGTEVYASGVWGTCSHMDSDASVICHOLQIGKGIAKQTFPSGIGL 175

```


QY 176 LIPIYW-SNVRCRDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 4
ID US-08-453-117-2 STANDARD: PRT: 451 AA.

XX xxxxxx

DE Sequence 2, Application US/08453117

CC Sequence 2, Application US/08453117

CC Patent No. 5683903

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Smithkline Beecham - Corporate Patents

CC ADDRESSEE: U.S.

CC STREET: Mailcode - UW2220, 709 Swedeland Road

CC CITY: King of Prussia

CC STATE: Pennsylvania

CC COUNTRY: U.S.A.

CC ZIP: 19406-5090

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/453,117

CC FILING DATE:

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Jervais, Herbert H.

CC REGISTRATION NUMBER: 31,171

CC REFERENCE/DOCKET NUMBER: SBC-P50338

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (610) 270-5019

CC TELEFAX: (610) 270-5090

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 451 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;

Query Match 31.0%; Score 237; DB 1; Length 451;

Best Local Similarity 37.9%; Pred. No. 9,70e-15; Indels 4; Gaps 4;

Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHRGREYELHSGQWGTICDDREVRVGVGVCSLGYPGVQAVHKAHF-GQG 408

QY 117 VRLGKNEFEGETVEYVASSVWGVCSHWDSDASVICHQLDGLG-KGIKQTPFSGLG 175

DB 409 TGPI-WLNEVFCFGRESSIECKIRQWGTACSHSEDAVYCT 450

QY 176 LIPIYW-SNVRCRDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 5
ID US-08-948-222-2 STANDARD: PRT: 451 AA.

XX xxxxxx

DE Sequence 2, Application US/08948222

XX Sequence 2, Application US/08948222

CC Patent No. 5683798

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells

CC NUMBER OF SEQUENCES: 4

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Smithkline Beecham - Corporate Patents

CC ADDRESSEE: U.S.

CC STREET: Mailcode - UW2220, 709 Swedeland Road

CC CITY: King of Prussia

CC STATE: Pennsylvania

CC COUNTRY: U.S.A.

CC ZIP: 19406-5090

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/948,222

CC FILING DATE:

CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US/08/453,117

CC FILING DATE:

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Jervais, Herbert H.

CC REGISTRATION NUMBER: 31,171

CC REFERENCE/DOCKET NUMBER: SBC-P50338

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (610) 270-5019

CC TELEFAX: (610) 270-5090

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 451 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;

Query Match 31.0%; Score 237; DB 2; Length 451;

Best Local Similarity 37.9%; Pred. No. 9,70e-15; Indels 4; Gaps 4;

Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHRGREYELHSGQWGTICDDREVRVGVGVCSLGYPGVQAVHKAHF-GQG 408

QY 117 VRLGKNEFEGETVEYVASSVWGVCSHWDSDASVICHQLDGLG-KGIKQTPFSGLG 175

DB 409 TGPI-WLNEVFCFGRESSIECKIRQWGTACSHSEDAVYCT 450

QY 176 LIPIYW-SNVRCRDEENILLCERDIWOGVCPQKMAAAYTCS 217

RESULT 6
ID US-08-973-145-2 STANDARD: PRT: 451 AA.

XX xxxxxx

DE Sequence 2, Application US/08973145

CC Sequence 2, Application US/08973145

CC Patent No. 5919636

CC GENERAL INFORMATION:

CC APPLICANT: Lysko, Paul G.

CC APPLICANT: Elshourbagy, Nabil A.

CC APPLICANT: Brawner, Mary E.

CC TITLE OF INVENTION: Attachment Enhanced 293 Cells


```

SC          LENGTH: 453
SQ          SEQUENCE 491 AA; 54317 MW; 1350575 CN;

Query Match          29.3%; Score 224; DB 4; Length 453;
Best Local Similarity 36.9%; Pred. No. 2,13e-13;
Matches 38; Conservative 18; Mismatches 43; Indels 4

Db 352 VALVAGSGPHERNVEFIHFEGCGMTGCTCDDRWELRGLVYCRSLGKRGVSYKKRAY
    ||| ||| : ||| ||| : ||| : ||| :
QY 117 VRLRGKNEFESTVAVYSGWGVYCSSHMDSDASVYCHQLQL-GKGIAKQTH
    ||| : ||| : ||| : ||| : ||| : ||| :
Db 411 TGPI-WLNEVEFCFPRESSIEECRIKMGVRCASHDEDAVGTCT 452
    ||| : ||| : ||| : ||| : ||| : ||| :
QY 176 LPIPIW-SNVRGRGDEINLLCEKNDIMGGVCPQMAAIVICS 217
    ||| : ||| : ||| : ||| : ||| : ||| :

T 9
US-08-316-714-10 STANDARD; PRT; 585 AA.
AC xxxxxx
DT
DX
XX
DE
XX
XX
XX
Sequence 10, Application US/08316714
CC Sequence 10, Application US/08316714
CC Patent No. 5965382
CC GENERAL INFORMATION:
CC APPLICANT: Koths, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casidyl, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Filthy Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA; 65330 MW; 1874129 CN;

Query Match          28.7%; Score 219; DB 2; Length 585;
Best Local Similarity 35.9%; Pred. No. 6.96e-13;

```

	MATCHES	37;	Conservative	21;	Mismatches	41;	Indels	4;	Gaps	4;
D6	24	MLRLDGAATNGRVEIFPRGQMGVCVCDNLMDLTASAVCRALGFENATOLGRAAF-GQG	82							
OY	117	VRLRGKNNEFEGTEYYVASGWYGVSHPDDSPASYICHOLOGKGRIA-KOTPFSLG	175							
D6	83	SGPILDEVOCTGTGEASLADC-KSLGMKSKCRHEDAGVCT	124							
OY	176	LIPITWSNVRRCRDENILLCEKDI-MOGGVCPQRMAAAVTCS	217							
XX	RESULT	10								
XX	ID	US-08-477-674-10	STANDARD:	PRT:	585	AA.				
DT	AC	xxxxxx								
DE	Sequence	10,	Application	US/08477674						
CC	Patent	No. 5644035								
CC	GENERAL INFORMATION:									
CC	APPLICANT:	Kotbs, Kirston E.								
CC	APPLICANT:	Halenbeck, Robert F.								
CC	APPLICANT:	Taylor, Eric W.								
CC	APPLICANT:	Wang, Alice M.								
CC	APPLICANT:	Caspihl, Clayton L.								
CC	TITLE OF INVENTION:	Secreted Mac-2-Binding Glycoprotein								
CC	NUMBER OF SEQUENCES:	11								
CC	CORRESPONDENCE ADDRESS:									
CC	ADDRESSEE:	Cetus Oncology Corporation								
CC	STREET:	1400 Fifty-Third Street								
CC	CITY:	Emeryville								
CC	STATE:	CA								
CC	COUNTRY:	USA								
CC	ZIP:	94608								
CC	COMPUTER READABLE FORM:									
CC	MEDIUM TYPE:	Floppy disk								
CC	COMPUTER:	IBM PC compatible								
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS								
CC	SOFTWARE:	Patentin Release #1.0, Version #1.25								
CC	CURRENT APPLICATION DATA:									
CC	APPLICATION NUMBER:	US/08/477,674								
CC	FILING DATE:	07-JUN-1995								
CC	CLASSIFICATION:	530								
CC	PRIOR APPLICATION DATA:									
CC	APPLICATION NUMBER:	US/08/316,714								
CC	FILING DATE:									
CC	APPLICATION NUMBER:	US/07/961,404								
CC	FILING DATE:	15-OCT-1992								
CC	ATTORNEY/AGENT INFORMATION:									
CC	NAME:	Goldman, Kenneth M.								
CC	REGISTRATION NUMBER:	34,174								
CC	REFERENCE/DOCKET NUMBER:	2595.1								
CC	TELECOMMUNICATION INFORMATION:									
CC	TELEPHONE:	(510) 420-3152								
CC	TELEFAX:	(510) 658-5470								
CC	TELEX:	N/A								
CC	INFORMATION FOR SEQ ID NO:	10:								
CC	SEQUENCE CHARACTERISTICS:									
CC	LENGTH:	585 amino acids								
CC	TYPE:	amino acid								
CC	STRANDEDNESS:	single								
CC	TOPOLOGY:	linear								
CC	MOLECULE TYPE:	protein								
SQ	SEQUENCE	585 AA; 65350 MW; 1874129 CN;								
Query Match		28.7%; Score 219; DB 1; Length 585;								
Best Local Similarity		35.9%; Pred. No. 6,96e-13;								
Matches	37; Conservative	21; Mismatches 41; Indels 4; Gaps 4;								
D6	24	MLRLDGAATNGRVEIFPRGQMGTCVNLMDLTASVCRALGFENATOLGRAAF-GQG	82							

OY 117 VALRGGKNEFEVGEVYVYASVWGTVCSSHHDDSDASVICHQLOGKGLA-KOTPFSGLG 175
DB 83 SCPIMDEVQCTGTEASLADC-KSLGWLKSNCRHERDAGVCT 124
OY 176 LPIYWSNVRCRGDEENILLCEKDI-WGGVCPQKMAAAVTC 217

RESULT 11
ID US-08-473-791-10 STANDARD; PRT; 585 AA.
AC xxxxxx
XX
XX
XX
XX

Sequence 10, Application US/08473791
DE Sequence 10, Application US/08473791
XX Patent No. 5736340
CC GENERAL INFORMATION:
CC APPLICANT: Kolts, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Caspitt, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-Third Street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/473,791
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA: 65330 MW: 1874129 CN;

Query Match 28.7%; Score 219; DB 1; Length 585;
Best Local Similarity 35.9%; Pred. No. 6.96e-13;
Matches 37; Conservative 21; Mismatches 41; Indels 4; Gaps 4;

DB 24 MRLAGATNCRGVEIFGYGOWGTCDNLMDLTDASVCRALGFENATQALGRAAF-GOG 82
OY 117 VALRGGKNEFEVGEVYVYASVWGTVCSSHHDDSDASVICHQLOGKGLA-KOTPFSGLG 175

DB 83 SCPIMDEVQCTGTEASLADC-KSLGWLKSNCRHERDAGVCT 124
OY 176 LPIYWSNVRCRGDEENILLCEKDI-WGGVCPQKMAAAVTC 217

RESULT 12
ID US-08-794-795-2 STANDARD; PRT; 495 AA.
AC xxxxxx
XX
XX
XX
XX

Sequence 2, Application US/08794795
DE Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourbagy, Nabli
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA: 49764 MW: 1218428 CN;

Query Match 28.4%; Score 217; DB 2; Length 495;
Best Local Similarity 41.2%; Pred. No. 1.12e-12;
Matches 42; Conservative 15; Mismatches 38; Indels 7; Gaps 5;

DB 399 VRIYSSNR-GRVYVYSGTWGTCDEDMONSDAIVFCRMLGYS-KGRALYKVGAGTG- 454
OY 117 VALRGGKNEFEVGEVYVYASVWGTVCSSHHDDSDASVICHQLOGKGLA-KOTPFSGLG 176
DB 455 -OI-MLDNVQCRGTESTLMSCTKNSGWHHDSHEEDAGVCS 494
OY 177 LPIYWSNVRCRGDEENILLCEKDIWGGVCPQKMAAAVTC 217

RESULT 13
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 6, Application US/08794795
CC
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourig, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Lysko, Paul
CC APPLICANT: Gross, Mitchell
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA; 52658 MW; 1340662 CN;
DB 424 VRIVGSSNR-GRAEVYISGTWGTICDDDEMONSDAIVFCMLGYS-KGRALIVYGAGTG-479
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQDLGGIKAKTPFSGGL 176
DB 480 -QI-WLDNVQCRGTESTLMSCTKNSMGHDCSHEDAGVICS 519
DY 177 IPIYV-SNVRGRDEENILLCERKDIMQGVCPKMAAAVICS 217
RESULT 14
ID US-08-470-350B-2 STANDARD: PRT: 1290 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 2, Application US/08470350B
CC
CC Sequence 2, Application US/08470350B
CC Patent No. 5664126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H
CC TITLE OF INVENTION: Enderin: A Secreted von Ebner's Gland
CC TITLE OF INVENTION: Protein Associated with Taste Buds
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107.48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1290 AA; 141879 MW; 9315616 CN;
DB 145 VRLVNGDRRCGRVETILLYGSGWGTCDSDINDANVCRQJGCGWALSAPSGAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQDLG-GKGIKAKTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYVSNVRGRDEENILLCERKDIMQGVCPKMAAAVICS 217
Query Match 28.4%; Score 217; DB 2; Length 520;
Best Local Similarity 41.2%; Pred. No. 1,12e-12;
Matches 42; Conservative 15; Mismatches 38; Indels 7; Gaps 5;

DT
XX
DE
XX
Sequence 2, Application US/08470350B
CC
XX
Sequence 2, Application US/08470350B
CC Patent No. 5664126
CC GENERAL INFORMATION:
CC APPLICANT: Li, Xiao
CC APPLICANT: Snyder, Solomon H
CC TITLE OF INVENTION: Enderin: A Secreted von Ebner's Gland
CC TITLE OF INVENTION: Protein Associated with Taste Buds
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1001 G Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,350B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wolfe, Susan A
CC REGISTRATION NUMBER: 33,568
CC REFERENCE/DOCKET NUMBER: 01107.48790
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-508-9100
CC TELEFAX: 202-508-9299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1290 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1290 AA; 141879 MW; 9315616 CN;
DB 145 VRLVNGDRRCGRVETILLYGSGWGTCDSDINDANVCRQJGCGWALSAPSGAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQDLG-GKGIKAKTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYVSNVRGRDEENILLCERKDIMQGVCPKMAAAVICS 217
Query Match 27.7%; Score 212; DB 1; Length 1290;
Best Local Similarity 37.3%; Pred. No. 3.63e-12;
Matches 38; Conservative 16; Mismatches 46; Indels 2; Gaps 2;
DB 145 VRLVNGDRRCGRVETILLYGSGWGTCDSDINDANVCRQJGCGWALSAPSGAQF-GOG 203
DY 117 VRLRGKNEFEETVEYVAGSWGTVCSHWDSDASYICHQDLG-GKGIKAKTPFSGGL 175
DB 204 SCSIVLADVACRGHEAVLWCSHRGWLSHNCGHQEDAGVICS 245
DY 176 IPIYVSNVRGRDEENILLCERKDIMQGVCPKMAAAVICS 217
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 2, Application US/08200900A
CC
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

Thu Mar 16 07:57:48 2000

US-09-147-947-6-03.ra1

Page 8

```

CC STREET: 87 Cambridgepark Drive
CC City: Cambridge
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02140
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/200,900A
CC FILING DATE: 23-FEB-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meinerdt, Maureen C.
CC REGISTRATION NUMBER: 31,544
CC REFERENCE/DOCKET NUMBER: GI 5201-FWC
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 876-1170 X8574
CC TELEFAX: (617) 876-5851
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 798 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 798 AA; 88937 MW; 3292434 CN;
Query Match 14.8%; Score 113; DB 1; Length 798;
Best Local Similarity 28.6%; Pred. No. 1,93e-02;
Matches 16; Conservative 10; Mismatches 29; Indels 1; Gaps 1;
Ddb 457 VRLNGTDSGGLVQFRRIQSIHWACAEWMTQTSDPYCOLLIGG-TGNSSVPYFS 511
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
117 VRLKGGKMEFEICIVLVASGVWGTVCSHWDSPDAISVICHOLQLGGKGIACTPES 172

```

Search completed: Mon Mar 13 10:26:11 2000
Job time : 8 secs.

117-217/5KR1

use k113 243

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Run on: Mon Mar 13 10:24:02 2000; Maspar time 6.39 Seconds
 Tabular output not generated. 472.013 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (117-217) from US09147947A.pep (3 of 6)
 Perfect Score: 764
 Sequence: 1 VRLRGKNEFEGTEVYASG.....KDIWGGVCPQKMAAVTCS 101

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: Swiss-Prot38
 1:swissprot

Statistics: Mean 38.635; Variance 55.528; scale 0.696

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
1	764	100.0	875	1 NTR_HUMAN	1.53e-182
2	323	42.3	761	1 NTR_MOUSE	6.23e-19
3	266	33.5	454	1 MSRE_RABIT	2.76e-41
4	244	31.9	458	1 MSRE_MOUSE	3.40e-38
5	237	31.0	451	1 MSRE_HUMAN	2.10e-36
6	236	30.5	1436	1 WC11_BOVIN	3.78e-36
7	233	30.9	453	1 MSRE_BOVIN	2.19e-35
8	193	25.3	532	1 SPER_STRPU	2.13e-25
9	188	24.6	468	1 CD6_HUMAN	3.52e-24
10	114	14.9	1069	1 ENTK_MOUSE	2.27e-07
11	113	14.8	1035	1 ENTK_BOVIN	3.64e-07
12	107	14.0	1034	1 ENTK_PIG	5.99e-06
13	105	13.7	1019	1 ENTK_HUMAN	1.48e-05
14	102	13.4	583	1 CFAT_HUMAN	5.71e-05
15	98	12.8	495	1 CD5_BOVIN	3.35e-04
16	91	11.9	494	1 CD5_MOUSE	6.71e-03
17	90	11.8	326	1 PEL_EMENI	1.02e-02
18	88	11.5	310	1 LMA2_HUMAN	2.32e-02
19	87	11.4	491	1 HMG2_YEAST	3.49e-02
20	84	11.0	495	1 CD5_HUMAN	1.16e-01
21	83	10.9	558	1 TF65_HUMAN	1.77e-01
22	81	10.6	1857	1 FAS2_PENPA	3.76e-01
23	80	10.5	204	1 LEF2_NEVOP	5.52e-01

HEX_ADEB3	HEXON PROTEIN (LATE PR	5.52e-01
24	80	10.5
25	79	10.3
26	78	10.2
27	78	10.2
28	77	10.1
29	77	10.1
30	77	10.1
31	76	9.9
32	76	9.9
33	76	9.9
34	76	9.9
35	76	9.9
36	76	9.9
37	75	9.8
38	75	9.8
39	75	9.8
40	75	9.8
41	74	9.7
42	74	9.7
43	74	9.7
44	74	9.7
45	74	9.7

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	875 AA.
ID	1			
AC	15-DEC-1999 (Rel. 39, Created)			
DT	15-DEC-1999 (Rel. 39, Last sequence update)			
DE	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	NEUTROTROPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).			
GN	PRSS12			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN:			
RA	MEDLINE: 98201705.			
RT	PROBA K., GSCHWEND T.P., SONDEREGGER P.,			
RL	"Cloning and sequencing of the cDNA encoding human neutrotropin.";			
CC	Biochim. Biophys. Acta 1396:143-147(1998).			
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC			
CC	ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH			
CC	LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL: AJ001531; CAA04816.1; -			
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.			
DR	PROSITE: PS00135; TRYPSIN_SER. 1.			
DR	PROSITE: PS00420; SPERACT_RECEPTOR. 3.			
KM	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	875	NEUTROTROPIN.
FT	DOMAIN	23	92	PROLINE-RICH.
FT	DOMAIN	93	165	KRINGLE.
FT	DOMAIN	170	271	SRCR 1.
FT	DOMAIN	170	280	SRCR 2.
FT	DOMAIN	387	487	SRCR 3.

```

FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 630 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 626 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA; 97011 MW; 67D5272B CRC32;

Query Match 100.0%; Score 764; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.53e-182;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 170 VRLHGKNEFGTEVYASVGMGTVCSSHMDDSDASVICHOLQGGKIAKOTPFSGGL 229
117 VRLHGKNEFGTEVYASVGMGTVCSSHMDDSDASVICHOLQGGKIAKOTPFSGGL 176
230 IPIYMSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 270
177 IPIYMSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 217

RESULT 2
ID NETR_MOUSE STANDARD: PRT: 761 AA.
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN) (BRAIN-SPECIFIC SERINE
GN PROTEASE 3) (BSSP-3).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE: 97401523.
RA GSCHEMEND T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neutrotypin, a novel multidomain serine protease expressed in the
RL nervous system.";
RN Mol. Cell. Neurosci. 9:207-219(1997).
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSURUOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RL "Molecular cloning of a novel brain-specific serine protease with a
RL trypsin-like structure and three scavenger receptor cysteine-rich
RL motifs.";
RN Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
CC LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
CC AND AMYGDALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Y13192; CAA73646.1; -.
DR EMBL: D89871; BAA23986.1; -.
DR MGD: MGI:1100881; PRSS12.
DR PFAM: PF00530; SRCR 3.
DR PFAM: PF00089; trypsin_1.
DR PROSITE: PS00134; TRYPSIN_HIS_1.
DR PROSITE: PS00135; TRYPSIN_SER_1.
DR PROSITE: PS00420; SPERACT_RECEPTOR_3.
DR Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 21
FT CHAIN 22 761 NEUTROTYPIN.
FT DOMAIN 85 157 KRINGLE.
FT DOMAIN 166 267 SRCR 1.
FT DOMAIN 273 373 SRCR 2.
FT DOMAIN 386 487 SRCR 3.
FT DOMAIN 505 761 SERINE PROTEASE.
FT DOMAIN 505 516 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 516 517 CHARGE RELAY SYSTEM.
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 POTENTIAL.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA; 84118 MW; 3F3C4F35 CRC32;

Query Match 42.3%; Score 323; DB 1; Length 761;
Best Local Similarity 45.5%; Pred. No. 6.23e-59;
Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Db 166 IRLVGSNGHEGRVLYAHAGWGTCIDQDMADADYICRQLGSGIAKAWQAHF-GEG 224
117 VRLHGKNEFGTEVYASVGMGTVCSSHMDDSDASVICHOLQGGKIA-KOTPFSGGL 175
225 SCPIILDEVRCTGNLSTFQCPKSSWGCHNGKEDAGVSC 265
176 IPIYMSNVRGDEENILLCERDIWOGVCPOKMAAAVTC 216

RESULT 3
ID MSRE_RABBIT STANDARD: PRT: 454 AA.
AC 005585;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
GN MSRL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93016877.
RA BICKEL P.E., FREEMAN M.W.;
RT "Rabbit aortic smooth muscle cells express inducible macrophage
RT scavenger receptor messenger RNA that is absent from endothelial
RT cells.";
RN J. Clin. Invest. 90:1450-1457(1992).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93131972.
RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
RA MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KAWABE Y.,
RA IMAKURA H., YAZAKI Y., MATSUMOTO A.;
RT "Charged collagen structure mediates the recognition of negatively
RT charged macromolecules by macrophage scavenger receptors.";
RL J. Biol. Chem. 268:2126-2133(1993).
CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE
CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING

```


CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L11693; AAA31402.1; -;
 CC EMBL: L11692; AAA31403.1; -;
 CC EMBL: D13381; BAA02648.1; -;
 CC PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC PFM: PFO0530; SRCR; 1.
 CC DR PFM: PFO1391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; Coiled coil; LDL;
 CC Heptad repeat pattern; Receptor; Alternative splicing.
 CC FT DOMAIN 1 50
 CC TRANSMEM 51 73
 CC FT DOMAIN 74 454
 CC FT DOMAIN 74 109
 CC FT DOMAIN 110 272
 CC FT DOMAIN 273 344
 CC FT DOMAIN 345 454
 CC FT CARBOHYD 82 82
 CC FT CARBOHYD 102 102
 CC FT CARBOHYD 143 143
 CC FT CARBOHYD 184 184
 CC FT CARBOHYD 221 221
 CC FT CARBOHYD 249 249
 CC FT CARBOHYD 267 267
 CC FT VARSPLIC 348 354
 CC FT VARSPLIC 355 454
 CC FT CONFLICT 106 106
 CC FT SEQUENCE 454 AA: 49745 MW: 5D780348 CRC32:
 CC
 CC Query Match 33.5%; Score 256; DB 1; Length 454;
 CC Best Local Similarity 40.8%; Pred. No. 2,76e-41;
 CC Matches 42; Conservative 17; Mismatches 40; Indels 4; Gaps 4;
 CC
 CC 353 VRLVGRGPHREGVEILHNGOMGTVCDDHVELRAGOVYCSLGIRGVKSYHKAYF-GQG 411
 CC 117 VRLRGKNEDEGEVEYVYASGVGTSVSHWDSDASVICHQLDL-GGKGIKOTPFESGLG 175
 CC DB 412 TGPI-WLNEVPCLGMESSECKIRQGVAVCSHGEDAGYTC 453
 CC QY 176 LRIPIY-SNVRCRGDEENILLCEKDIHGGVCPQKMAAAVTC 217
 CC
 CC RESULT 4
 CC ID MSRE.MOUSE STANDARD: PRT: 458 AA.
 CC AC P30204;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-DEC-1999 (Rel. 39, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC LDL RECEPTOR I AND II).
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 93359822.
 CC RA ASHKENAS J., PENMAN M., VASILE E., ACTON S., FREEMAN M.W.,
 CC KRIEGER M.;

RT "Structures and high and low affinity ligand binding properties of
 RT murine type I and type II macrophage scavenger receptors.";
 RT J. Lipid Res. 34:983-1000(1993).
 RL [2]
 RL RP SEQUENCE FROM N.A. (SHORT FORM).
 RX MEDLINE: 93331972.
 RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
 RA MIYAKAKI T., NAKAMURA H., DESUGI S., IMANISHI T., KAMAE Y.,
 RA ITAKURA H., YAZAKI Y., MATSUMOTO A.;
 RA "Charged collagen structure mediates the recognition of negatively
 RT charged macromolecules by macrophage scavenger receptors.";
 RL J. Biol. Chem. 268:2126-2133(1993).
 RN [3]
 RN RP SEQUENCE OF 349-458 FROM N.A.
 RX MEDLINE: 91062370.
 RA FREEMAN M., ASHKENAS J., REES D.J., KINGSLEY D.M., COPELAND N.G.,
 RA JENKINS N.A., KRIEGER M.;
 RA "An ancient, highly conserved family of cysteine-rich protein domains
 RT revealed by cloning type I and type II murine macrophage scavenger
 RT receptors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).
 RN [4]
 RN RP SEQUENCE OF 1-4 FROM N.A.
 RX MEDLINE: 95395388.
 RA AETRING R.P., FREEMAN M.W.;
 RA "Structure of the murine macrophage scavenger receptor gene and
 RT evaluation of sequences that regulate expression in the macrophage
 RT cell line, P388D.";
 RL J. Lipid Res. 36:1305-1314(1995).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIANE THE
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: I04274; AAA39747.1; ALT_INIT.
 CC EMBL: I04275; AAA39748.1; ALT_INIT.
 CC EMBL: D13382; BAA02650.1; -;
 CC EMBL: M59445; AAA37464.1; -;
 CC EMBL: M59446; AAA37465.1; -;
 CC EMBL: U13873; AAC13774.1; -;
 CC PIR: A38260; A38260.
 CC MGI: 98257; SCVR.
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFM: PFO0530; SRCR; 1.
 CC DR PFM: PFO1391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 CC Heptad repeat pattern; Receptor; Alternative splicing.
 CC FT DOMAIN 1 55
 CC TRANSMEM 56 78
 CC FT DOMAIN 79 458
 CC FT DOMAIN 79 114
 CC FT DOMAIN 114 276
 CC FT DOMAIN 277 350
 CC FT DOMAIN 351 458
 CC FT CARBOHYD 94 94
 CC FT CARBOHYD 107 107
 CC FT CARBOHYD 147 147
 CC FT CARBOHYD 188 188
 CC
 CC EXTRACELLULAR (POTENTIAL).
 CC SPACER (PROBABLE).
 CC ALPHA-HELICAL COILED-COIL.
 CC COLLAGEN-LIKE.
 CC CYS-RICH.
 CC POTENTIAL.
 CC POTENTIAL.
 CC POTENTIAL.
 CC POTENTIAL.

FT CARBOHYD 253 253 POTENTIAL.
 FT CARBOHYD 271 271 POTENTIAL.
 FT VARSPLIC 352 354 TPL -> RSV (IN ISOFORM II).
 FT VARSPLIC 355 458 MISSING (IN ISOFORM II).
 SQ SEQUENCE 458 AA; 50130 MW; F28A456E CRC32;

Query Match 31.9%; Score 244; DB 1; Length 458;
 Best Local Similarity 36.9%; Pred. No. 3,40e-38;
 Matches 38; Conservative 19; Mismatches 42; Indels 4; Gaps 4;

DB 357 VRLVGGSGAHEGRVEITHQGMGTICDDRMIDRAGVVCRSIGYOEVAHKAHF-GOG 415
 117 VRLRGKNEEGVEYVAGVWGVCSSHWDSDASYICHQLDGG-GKIAKOTPFSGLG 175

DB 416 TGP1-WLNEVFCFGRESSIENCKINQWGVLSGSHSDEGATCT 457
 176 LIPIYW-SNVRCRDEENILCEKIDWOGVCPQKMAAVTCS 217

5
 ASRE_HUMAN STANDARD; PRT; 451 AA.
 P21757; P21759;
 01-MAY-1991 (Rel. 18, Created)
 01-MAY-1991 (Rel. 18, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 LDL RECEPTOR I AND II).
 GN MSRI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE; 91067661.
 RA MATSUMOTO A., NAITO M., ITAKURA H., IKEMOTO S., ASAKA H.,
 HAYAKAWA I., ABURATANI H., TAKAKU F., SUZUKI H.,
 KOBARI Y., MIYAI T., TAKAHASHI K., COHEN E.H., WIDRO R.,
 HOSMAN D.E., KODAMA T.;
 "Human macrophage scavenger receptors: primary structure, expression,
 and localization in atherosclerotic lesions."
 Proc. Natl. Acad. Sci. U.S.A. 87:9133-9137(1990).
 CC -! FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -! SUBUNIT: HOMOTRIMER.
 CC -! SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -! ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 OF THE SAME GENE.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; D90187; BAA14208.1; -
 CC EMBL; D90188; BAA14209.1; -
 DR PIR; A38415; A38415.
 DR PIR; B38415; B38415.
 DR MIM; 153622; -
 DR PROSITE; PS00420; SPERACT_RECEPTOR; 1.
 DR PFAM; PF00530; SRCR; 1.
 DR PFAM; PF01391; Collagen; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 Heptad repeat pattern; Receptor; Alternative splicing.
 FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 51 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 77 451 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 77 109 SPACER (PROBABLE).
 FT DOMAIN 110 272 ALPHA-HELICAL COILED-COIL.
 FT DOMAIN 273 341 COLLAGEN-LIKE.
 FT DOMAIN 342 451 CYS-RICH.
 FT CARBOHYD 82 82 POTENTIAL.
 FT CARBOHYD 102 102 POTENTIAL.
 FT CARBOHYD 143 143 POTENTIAL.
 FT CARBOHYD 184 184 POTENTIAL.
 FT CARBOHYD 221 221 POTENTIAL.
 FT CARBOHYD 249 249 POTENTIAL.
 FT CARBOHYD 267 267 POTENTIAL.
 FT VARSPLIC 345 358 POTENTIAL.
 FT VARSPLIC 359 451 ITPKRLVIGSGSP -> RPVLDTIRAGPS (IN
 ISOFORM II).
 SQ SEQUENCE 451 AA; 49762 MW; AB9FEED7 CRC32;

Query Match 31.0%; Score 237; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 2,10e-36;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHEGRVEILHSQMGITICDDREVEVGVVCRSLGYPGOVAKRAHF-GOG 408
 117 VRLRGKNEEGVEYVAGVWGVCSSHWDSDASYICHQLDGG-KGIAKOTPFSGLG 175

DB 409 TGP1-WLNEVFCFGRESSIECKIRQWGTACSHSDEGATCT 450
 176 LIPIYW-SNVRCRDEENILCEKIDWOGVCPQKMAAVTCS 217

RESULT 6
 ID WC11_BOVIN STANDARD; PRT; 1436 AA.
 AC P30205;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE ANTIGEN WC1.1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 CC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BLOOD.
 RX MEDLINE; 93056489.
 RA WUJINGARD P.L.J., METZELAR M.J., MACHUGH N.D., MORRISON W.I.,
 CLEVERS H.C.;
 RT "Molecular characterization of the WC1 antigen expressed specifically
 on bovine CD4-CD8-gamma delta T lymphocytes."
 RL J. Immunol. 149:3273-3277(1992).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; X63723; CAA45255.1; -
 DR PIR; S19913; S19913.
 DR PIR; A46496; A46496.
 DR PROSITE; PS00420; SPERACT_RECEPTOR; 4.
 DR PFAM; PF00530; SRCR; 1.
 KW Antigen; Repeat.
 SQ SEQUENCE 1436 AA; 154196 MW; D61545D7 CRC32;

Query Match 30.9%; Score 236; DB 1; Length 1436;
 Best Local Similarity 37.9%; Pred. No. 3,78e-36;
 Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;

DB 115 VRLRGSGSGRVEVHNSMGTCDDSDSLAEVVCQGLGCGALVAVSAAR-GPG 1213
 117 VRLRGKNEEGVEYVAGVWGVCSSHWDSDASYICHQLDGG-GKIAKOTPFSGLG 175

DB 1214 NSST-WLDEVQCGGREGSSIMDCVAEPWGOSDCKHEPDAGVCS 1255
 QY 176 LPIYW-SNVRCRGDENILLCEKDIMOGVCPQKMAAVTCS 217

RESULT 7 STANDARD: PRT: 453 AA.

ID MSRE_BOVIN
 AC P21758;
 DT 01-MAY-1991 (Rel. 18, Created)
 DI 01-MAY-1991 (Rel. 18, Last sequence update)
 DE 01-FEB-1994 (Rel. 28, Last annotation update)
 DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 LDL RECEPTOR I AND II).
 CS Bos taurus (Bovine).
 CS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 Bovinae; Bos.

[1]
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=LUNG;
 RX MEDLINE: 90136965.
 RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
 RA KRIEGER M.;
 RT "Type I macrophage scavenger receptor contains alpha-helical and
 RT collagen-like coiled coils.";
 RL Nature 343:531-535(1990).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=LUNG;
 RX MEDLINE: 90136973.
 RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIEGER M.;
 RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
 RT scavenger receptor type II.";
 RL Nature 343:570-572(1990).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIANE THE
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X51689; CA35987.1;
 DR EMBL: X54183; CA38108.1;
 DR PIR: S08276; S08276.
 DR PIR: S08278; S08278.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF01391; Collagen; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 KW Heptad repeat pattern; Receptor; Alternative splicing.
 FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 51 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 77 453 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 77 108 SPACER (PROBABLE).
 FT DOMAIN 109 271 ALPHA-HELICAL COILED-COIL.
 FT DOMAIN 272 343 COLLAGEN-LIKE.
 FT DOMAIN 344 453 CYS-RICH.
 FT CARBOHYD 82 82 POTENTIAL.
 FT CARBOHYD 101 101 POTENTIAL.
 FT CARBOHYD 142 142 POTENTIAL.

FT CARBOHYD 183 183 POTENTIAL.
 FT CARBOHYD 220 220 POTENTIAL.
 FT CARBOHYD 248 248 POTENTIAL.
 FT CARBOHYD 266 266 POTENTIAL.
 FT VARSPLIC 348 349 OS -> PG (IN ISOFORM II).
 FT VARSPLIC 350 453 MISSING (IN ISOFORM II).
 SQ SEQUENCE 453 AA; 50056 MW; 2CDEIEEI CRC32;

Query Match
 Best Local Similarity 30.5%; Score 233; DB 1; Length 453;
 Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

DB 352 VRLVGGSGPHEGRVEIFHEGQWTCDDRWELRGILVCRSLGYGVQSVHRRAYF-GKG 410
 QY 117 VRLRGKNEFEETVEYASGVGTCCSHWDDSDASVICHQLT-GGKGIKQTPSSIG 175

DB 411 TGPI-WLNVFPGKSSIEEERIRQWGRACSHDEDAVTC 452
 QY 176 LPIYW-SNVRCRGDENILLCEKDIMOGVCPQKMAAVTCS 217

RESULT 8 STANDARD: PRT: 532 AA.

ID SPER_SRRPU
 AC P16264;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EGG PEPTIDE SPERACT RECEPTOR PRECURSOR.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 OC Echinoidea; Echinoidea; Echinoidea; Strongylocentrotidae;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.
 RX MEDLINE: 89184581.
 RA DANCOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;
 RT "Cloning of the mRNA for the protein that crosslinks to the egg
 RT peptide speract.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).
 CC -1- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPERACT.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: J04518; AAA30078.1;
 DR PIR: A32751; A32751.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 4.
 DR PFAM: PF00530; SRCR; 4.
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
 FT SIGNAL 1 30
 FT CHAIN 31 532 EGG PEPTIDE SPERACT RECEPTOR.
 FT DOMAIN 31 491 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 492 520 POTENTIAL.
 FT DOMAIN 521 532 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 40 488 4 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 40 147 1.
 FT REPEAT 148 260 2.
 FT REPEAT 261 375 3.
 FT REPEAT 376 488 4.
 FT CARBOHYD 78 78 POTENTIAL.
 FT CARBOHYD 115 115 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 SQ SEQUENCE 532 AA; 57820 MW; D21774AE CRC32;

Query Match
 Best Local Similarity 25.3%; Score 193; DB 1; Length 532;
 Matches 38; Conservative 19; Mismatches 41; Indels 8; Gaps 6;

Db 264 IRMGSGPHEGRVEIMHDDANGTICDDGMDANAVYCRQ-A-GYRGVAKSGFGEDE 321
 QY 117 VRLRGKNEFEETVEYVYSGVGTWCSSHMDDSDASVICHQLOLGKGIKOPFSG--L 174
 Db 322 GFTWAPIHTEFVCTGVEDRLIDCIRLDGWTMS-CYHVDASVYCA 366
 QY 175 GLI--PIYMSNVRCRGDEENILCE-KDIMGGVCPQKMAAVTCS 217

RESULT 9
 ID CD6_HUMAN STANDARD: PRT: 468 AA.
 AC P30203:
 DI 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL DIFFERENTIATION ANTIGEN CD6 PRECURSOR (T12) (T120).
 GN CD6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RN Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92013796.
 RA ARUFFO A., MELNICK M.B., LINSLEY P.S., SEED B.:
 RT "The lymphocyte glycoprotein CD6 contains a repeated domain structure
 characteristic of a new family of cell surface and secreted
 proteins.";
 RL J. Exp. Med. 174:949-952(1991).
 CC -1- FUNCTION: INVOLVED IN CELL ADHESION. BINDS TO CD166.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD6 entry;
 WWW="http://www.ncbi.nlm.nih.gov/PROV/cd/cd6.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X60992; CAA43306.1; -
 DR PIR: S26741; S26741.
 DR MIM: 186720; -
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 DR PFM: PF00530; SRCR; 3.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal.
 FT SIGNAL 1
 FT MAIN ? 468 T-CELL DIFFERENTIATION ANTIGEN CD6.
 FT DOMAIN ? 402 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 403 423 POTENTIAL.
 FT DOMAIN 424 468 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 28 28 POTENTIAL.
 FT CARBOHYD 49 49 POTENTIAL.
 FT CARBOHYD 112 112 POTENTIAL.
 FT CARBOHYD 118 118 POTENTIAL.
 FT CARBOHYD 229 229 POTENTIAL.
 FT CARBOHYD 339 339 POTENTIAL.
 FT CARBOHYD 345 345 POTENTIAL.
 FT CARBOHYD 368 368 POTENTIAL.
 SQ SEQUENCE 468 AA; 49828 MW; 042D5AC8 CRC32;

Query Match 24.6%; Score 188; DB 1; Length 468;
 Best Local Similarity 36.6%; Pred. No. 3.52e-24;
 Matches 37; Conservative 19; Mismatches 33; Indels 6; Gaps 4;

Db 266 RLITGADRCGQVEVHFRGVNTVCDSEWYPSAEKVLQSLG-CGTAVERPGLPHSIS- 323
 QY 118 RLRGKNEFEETVEYVYSGVGTWCSSHMDDSDASVICHQLOLGKGIKOPFSG--L 176
 Db 324 GRWYYS---CNGEELTJLSCNMFNNSNLCSSSLAARVLCS 361

QY 177 IPYMSNVRCRGDEENILCEKDIMGGVCPQKMAAVTCS 217
 RESULT 10
 ID ENTK_MOUSE STANDARD: PRT: 1069 AA.
 AC P97435:
 DI 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 RN Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=DUODENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.:
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-ILS-1-ILE-7 BOND IN
 TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PIV: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDB-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U73378; AA37317.1; -
 DR HSBP; P00763; IDPO.
 DR MGD; MGI:1197523; PRSS7.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; TRYPSIN_SER; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS50209; IDLRA_1; 2.
 DR PROSITE: PS50068; IDLRA_2; 2.
 DR PFM: PF00057; Idl_recept_a; 2.
 DR PFM: PF00089; trypsin; 1.
 DR PFM: PF00431; CUB; 2.
 DR PFM: PF00530; SRCR; 1.
 DR PFM: PF00629; MAM; 1.
 DR PFM: PF01390; SEA; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 Setine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 227 268 IDL-RECEPTOR CLASS A 1.
 FT DOMAIN 270 379 CUB.
 FT DOMAIN 387 549 MAM.
 FT DOMAIN 569 679 CUB.

FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 723 816 SRCR.
 FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT LIPID 2 2 MYRISTATE (POTENTIAL).
 FT DISULFID 229 242 BY SIMILARITY.
 FT DISULFID 236 255 BY SIMILARITY.
 FT DISULFID 249 266 BY SIMILARITY.
 FT DISULFID 688 700 BY SIMILARITY.
 FT DISULFID 695 713 BY SIMILARITY.
 FT DISULFID 707 722 BY SIMILARITY.
 FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 859 875 BY SIMILARITY.
 FT DISULFID 959 1027 BY SIMILARITY.
 FT DISULFID 991 1006 BY SIMILARITY.
 FT DISULFID 1017 1045 BY SIMILARITY.
 FT CARBOHYD 147 147 POTENTIAL.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 212 212 POTENTIAL.
 FT CARBOHYD 373 373 POTENTIAL.
 FT CARBOHYD 380 380 POTENTIAL.
 FT CARBOHYD 433 433 POTENTIAL.
 FT CARBOHYD 515 515 POTENTIAL.
 FT CARBOHYD 579 579 POTENTIAL.
 FT CARBOHYD 675 675 POTENTIAL.
 FT CARBOHYD 727 727 POTENTIAL.
 FT CARBOHYD 751 751 POTENTIAL.
 FT CARBOHYD 770 770 POTENTIAL.
 FT CARBOHYD 791 791 POTENTIAL.
 FT CARBOHYD 897 897 POTENTIAL.
 FT CARBOHYD 936 936 POTENTIAL.
 FT CARBOHYD 999 999 POTENTIAL.
 SO SEQUENCE 1069 AA; 118735 MM; 48B825A2 CRC32;

Query Match 14.9%; Score 114; DB 1; Length 1069;
 Best Local Similarity 26.7%; Pred. No. 2.27e-07;
 Matches 16; Conservative 17; Mismatches 25; Indels 2; Gaps 2;

Db 723 VRFNGTRSN-NGIYQINISIMHACENMTQINSVCHGLTGSANSMPISSTGGG 781
 117 VR-LRGKNEFEQVETVYASGWTGVCSSHWSDASVICHOLQGGKIAKOTPFESGLG 175

RESULT 11 STANDARD: PRT: 1035 AA.
 ID ENTK BOVIN
 01-FEB-1996 (Rel. 33, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 15-DEC-1999 (Rel. 39, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PPS7 OR ENTK.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 CC Bovinae; Bos.
 [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-DUODENUM.
 RX MEDLINE: 94329561.
 RA KILAMOTO Y., YUAN X., WU O., MCCOURT D.W., SADLER J.E.;
 "Enterokinase, the initiator of intestinal digestion, is a mosaic
 protease composed of a distinctive assortment of domains.";
 Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 [2]
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 94043122.
 RA LAVALLE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
 FERENZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
 "Cloning and functional expression of a cDNA encoding the catalytic
 subunit of bovine enterokinase.";
 J. Biol. Chem. 268:23311-23317(1993).
 [3]

RP SEQUENCE OF 801-827.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 92189715.
 RA LIGHT A., JANSKA H.;
 "The amino-terminal sequence of the catalytic subunit of bovine
 enterokinase.";
 J. Protein Chem. 10:475-480(1991).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: U09859; AAB40026.1;
 CC EMBL: L19663; AAL16035.1;
 CC PIR: A61436; A61436.
 CC HSSP: P00763; IDPO.
 DR PROSITE: PS00134; TRYPSIN_HTS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 DR PFAM: PF00057; ldl_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00431; CUB; 2.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF00629; MAM; 1.
 DR PFAM: PF01390; SEA; 1.
 KW Signal anchor; Glycoprotein; Myristate; Hydroxylase;
 KW Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 240 350 CUB.
 FT DOMAIN 358 520 MAM.
 FT DOMAIN 540 650 CUB.
 FT DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 694 787 SRCR.
 FT ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 882 882 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 987 987 MYRISTATE (POTENTIAL).
 FT LIPID 2 2 MYRISTATE (POTENTIAL).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 219 236 BY SIMILARITY.
 FT DISULFID 659 671 BY SIMILARITY.
 FT DISULFID 666 684 BY SIMILARITY.
 FT DISULFID 678 693 BY SIMILARITY.

FT CARBOHYD 804 804 POTENTIAL.
 FT CARBOHYD 863 863 POTENTIAL.
 FT CARBOHYD 902 902 POTENTIAL.
 FT CARBOHYD 964 964 POTENTIAL.
 SO SEQUENCE 1034 AA; 114776 MW; 24386471 CRC32;

Query Match 14.0%; Score 107; DB 1; Length 1034;
 Best Local Similarity 28.6%; Pred. No. 5,93e-06;
 Matches 17; Conservative 12; Mismatches 29; Indels 1; Gaps 1;

Db 693 VRLNGTANNGLVGRFRIOSIWHITACAEWNTQTSDDVQCLLIG-TGNSSMPFFSSGG 750
 117 VRLRGKNEFEETVEYVAGSVGTWCSSHWDDSDASYICHQLDGLGKGIKOTPPSGLG 175

Query Match 13
 ENT_K_HUMAN STANDARD; PRT: 1019 AA.
 998073;
 01-FEB-1996 (Rel. 33, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRC57 OR ENT_K.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=DUODENUM;
 RX MEDLINE: 95234679.
 RA KITAMOTO Y., VEILE R.A., DONIS-KELLER H., SADLER J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 the proteolytic activator of trypsinogen.";
 PL Biochemistry 34:4562-4568(1995).
 [2]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=DUODENUM;
 RX MEDLINE: 94329561.
 KA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 protease composed of a distinctive assortment of domains.";
 PL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
 TRYPsinOGEN.
 - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 - TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 - DISEASE: DEFECTS IN PRC57 CAUSE LIFE-THREATENING INTESTINAL
 MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
 - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 - This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: U09860; AAC50138.1;

DR HSP; P00763; IDPO.
 DR MIM: 226200; -;
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS00060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 DR PFAM: PF00057; LDL_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00431; CUB; 2.
 DR PFAM: PF00530; SRCR; 1.
 DR PFAM: PF00629; MAM; 1.
 DR PFAM: PF01390; SEA; 1.
 DR Signal-anchor: Glycoprotein: Myristate; Hydrolase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 784
 FT TRANSMEM 19 47
 FT DOMAIN 182 223
 FT DOMAIN 225 334
 FT DOMAIN 342 504
 FT DOMAIN 524 634
 FT DOMAIN 641 679
 FT ACT_SITE 825 825
 FT ACT_SITE 876 876
 FT ACT_SITE 971 971
 FT LIPID 2
 FT DISULFID 184 197
 FT DISULFID 191 210
 FT DISULFID 204 221
 FT DISULFID 643 655
 FT DISULFID 650 668
 FT DISULFID 662 677
 FT DISULFID 772 896
 FT DISULFID 910 977
 FT DISULFID 941 956
 FT DISULFID 967 995
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 179 179
 FT CARBOHYD 328 328
 FT CARBOHYD 335 335
 FT CARBOHYD 388 388
 FT CARBOHYD 440 440
 FT CARBOHYD 470 470
 FT CARBOHYD 503 503
 FT CARBOHYD 534 534
 FT CARBOHYD 630 630
 FT CARBOHYD 682 682
 FT CARBOHYD 706 706
 FT CARBOHYD 725 725
 FT CARBOHYD 848 848
 FT CARBOHYD 887 887
 FT CARBOHYD 909 909
 FT CARBOHYD 949 949
 SO SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;

Query Match 13.7%; Score 105; DB 1; Length 1019;
 Best Local Similarity 28.6%; Pred. No. 1.48e-05;
 Matches 16; Conservative 13; Mismatches 26; Indels 1; Gaps 1;

Db 678 VREFNGTNNGLVGRFRIOSIWHITACAEWNTQISNDVQCLLIGSGNSKPI-FS 732
 117 VRLRGKNEFEETVEYVAGSVGTWCSSHWDDSDASYICHQLDGLGKGIKOTPPS 172

Query Match 14
 ENT_K_HUMAN STANDARD; PRT: 583 AA.
 ID CFAI_HUMAN
 AC P05156;

SQ SEQUENCE 495 AA; 5433 MW; F8ED2AE0 CRC32;
 Query Match 12.8%; Score 98; DB 1; Length 495;
 Best Local Similarity 39.1%; Pred. No. 3.35e-04;
 Matches 18; Conservative 9; Mismatches 15; Indels 4; Gaps 4;
 Db 275 RLVGSDVCEGSEVFRSGKGKMDLDDSWAKGTARREVCREQQ 320
 QY 118 RLRGKNEFEETVEYAS-GV-WGTVCSSHMDSDAS-V-ICHOLQ 159

Search completed: Mon Mar 13 10:24:13 2000
 Job time : 11 secs.

THIS PAGE BLANK (USPTO)

117-217/SRCR21

Nothing better than PIRQ Swissprot butts

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Mon Mar 13 10:24:29 2000;      MasPar time 16.49 Seconds
Tabular output not generated.      424.656 Million cell updates/sec
```

```

Title: >US-09-147-947-6
Description: (117-217) from US09147947A-pep (3 of 6)
Perfect Score: 764
Sequence: 1 VRLGKGNKEDEGVEYVASC.....KINMGVCYCPQKMAANTYCS 101

```

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: spremb112

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_plage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 37.758; Variance 56.548; scale 0.666

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					Pred.	
No.	Score	Match	Length	DB	ID	Description	No.
1	266	34.8	2153	5	097375	SCAVENGER RECEPTOR CYS	4.51e-42
2	254	33.2	2043	5	096943	SRCA DOMAIN, MEMBRANE	4.51e-39
3	251	32.9	1036	5	097378	SCAVENGER RECEPTOR CYS	2.51e-38
4	251	32.9	1116	4	097988	M130 ANTIGEN PRECURSOR	2.51e-38
5	251	32.9	1149	4	007901	M130 ANTIGEN, EXTRACEL	2.51e-38
6	251	32.9	1151	4	007899	M130 ANTIGEN, CYTOPLAS	2.51e-38
7	251	32.7	1156	4	007900	M130 ANTIGEN, CYTOPLAS	2.51e-38
8	250	32.7	754	11	092175	LYSLYL OXIDASE-RELATED	4.45e-38
9	243	31.8	688	4	0945X8	LYSLYL OXIDASE-LIKE PRO	2.41e-36
10	243	31.8	774	4	0945K0	LYSLYL OXIDASE-RELATED	2.41e-36
11	242	31.7	480	6	028112	SCAVENGER-RECEPTOR PRO	4.26e-35
12	242	31.7	804	6	029113	SCAVENGER-RECEPTOR PRO	4.26e-35
13	240	31.4	1785	4	0294V9	DMB1/16KB, 1 PROTEIN PR	1.33e-35
14	240	31.4	1785	4	029211	DMB1 PROTEIN.	1.33e-35
15	239	31.3	369	6	029110	SCAVENGER-RECEPTOR PRO	2.34e-35
16	239	31.3	458	11	060754	BACTERIA BINDING MACRO	2.34e-35
17	238	31.2	427	6	029109	SCAVENGER-RECEPTOR PRO	4.12e-35
18	234	30.6	800	6	028810	MUCIN (FRAGMENT).	3.96e-34
19	230	30.5	1554	6	092518	HEN5IN.	6.97e-34
20	230	30.1	918	13	092098	PEAK-SRCR PROTEIN PREC	3.79e-33

21	225	29.5	347.4	0438663	SP-ALPHA.	6.29e-32
22	219	28.7	585.4	008380	MAC-2 BINDING PROTEIN	1.81e-30
23	217	28.4	483.11	09WUB9	MACROPHAGE RECEPTOR MA	5.51e-30
24	217	28.4	520.4	059533	MACROPHAGE RECEPTOR	5.51e-30
25	215	28.1	2083.11	060957	CCR-DUCTIN PRECURSOR	1.66e-29
26	212	27.7	356.6	097682	UNKNOWN MRNA, PARTIAL	8.87e-29
27	212	27.7	574.11	070513	UNKNOWN.	8.87e-29
28	212	27.7	1290.11	062827	EBERNRN.	8.87e-29
29	211	27.6	822.13	09HCL1	LYSYL OXIDASE HOMOLOG	1.54e-28
30	211	27.6	895.13	09W6M1	LYSYL OXIDASE RELATED	1.54e-28
31	210	27.5	868.5	09Y1V3	TUNICATE RETINOIC ACID	2.65e-28
32	209	27.4	127.6	095316	M30 ANTIGEN (FRAGMENT	4.67e-28
33	205	27.0	504.6	082908	MUCIN (FRAGMENT).	2.45e-27
34	205	26.8	578.11	P70117	PANCREAS CANCER ASSOCI	2.45e-27
35	204	26.7	665.11	061003	T-CELL DIFFERENTIATION	7.36e-27
36	202	26.4	531.5	110764	SEB5.	2.21e-26
37	202	26.4	577.11	007797	PEPTIDYLPROLYL ISOMERA	2.21e-26
38	198	25.9	532.5	025111	A SPERM-ACTIVATING PEP	1.97e-25
39	194	25.4	352.11	035301	SP-ALPHA.	1.75e-24
40	194	25.4	352.11	035300	SP-ALPHA.	1.75e-24
41	191	25.0	462.6	028881	MEMBRANE PROTEIN SCAYE	8.92e-24
42	191	25.0	574.11	035669	CYCLOPHILIN C-ASSOCIAT	8.92e-24
43	188	24.6	595.4	09Y4K9	CP6C.	4.52e-23
44	188	24.6	601.4	09Y4R8	CP6D.	4.52e-23
45	188	24.6	635.4	09Y4L0	CP6B.	4.52e-23

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	2153 AA.
ID	097375			
AC	097375;			
DT	01-MAY-1999 (TReMBLrel. 10, Created)			
DT	01-MAY-1999 (TReMBLrel. 10, Last sequence update)			
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)			
DE	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.			
GN	SCRC12.			
OS	Strongylocentrotus purpuratus (Purple sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidae;			
OC	Euechinozoa; Echinacea; Echinoida; Strongylocentrotidae;			
OC	Strongylocentrotus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	PANCER Z., RAST J., DAVIDSON E.H.;			
RT	"Evolution of the Immune System: Transcription Factors and Downstream			
RT	Genes of the Mammalian Immune System Expressed in Sea Urchin			
RT	Coelomocytes."			
RL	Immunogenetics 0:0-0(1999).			
DR	EMBL; AF064259; AAD08654.1; -;			
DR	PROSITE; PS00420; SPERACT_RECEPTOR; 15.			
FT	Signal; Receptor.			
FT	SIGNAL	1	16	POTENTIAL.
FT	CHAIN	17	2153	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN
FT				TYPE 12.
SO	SEQUENCE	2153 AA;	226557 MW;	CA3BA8A9 CRC32.
	Query Match	34.8%;	Score 266;	DB 5; Length 2153;
	Best Local Similarity	38.6%;	Pred. No. 4,54e-42;	
	Matches 39;	Conservative	22;	Mismatches 40; Indels 0; Gaps
Db	462 VRLVGGNNREGREIFELNNQMGVCDMDGTPRPNANYCRQLGSPGSGRSSAYGRGS	5211		
Qy	117 VRLGKGNEFGIYEVLAAGVGVGVCSSHHDDSDASYICHQLOLGCGIKAKIPFSGLGL	176		
Db	522 VILLDNVGCGSGNRSLELCSNNNGIGVHNCGHQEDASVCT	562		
Qy	177 IPIYMSNVRCRGDEENITLCEKIDIMQGVCPQKMAAAVTC	217		
RESULT	2	PRELIMINARY;	PRT;	2043 AA.
ID	096943			
AC	096943;			
DT	01-MAY-1999 (TReMBLrel. 10, Created)			

RA MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily."
RL Eur. J. Immunol. 23:2320-2325(1993).
CP EMBL: 222971; CA80544.1; -
PF PRAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1149
SQ SEQUENCE 1149 AA; 124328 MW; 4901C708 CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1149;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
957 IRLQEGTSCSGREIWHGSGMTGVCDDSDMDLDAQVCCOLGCGPALKAFAEF-GQG 1015
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
DB 1016 TGPI-WLNEVKCKGNESLMDPCPARWGSHCGHKEDAANCT 1057
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
QY 176 LIPIYW-SNVRGRDDENILLCEKDIMGVCPQKMAAAVTC5 217
RESULT 6
ID 007899 PRELIMINARY; PRT: 1151 AA.
AC 007899;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE M30 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA IAW S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C., MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily."
RL Eur. J. Immunol. 23:2320-2325(1993).
DR EMBL: 222969; CA80542.1; -
DR PRAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1151
SQ SEQUENCE 1151 AA; 124820 MW; A7ZEDD2F CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1151;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
924 IRLQEGTSCSGREIWHGSGMTGVCDDSDMDLDAQVCCOLGCGPALKAFAEF-GQG 982
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
DB 983 TGPI-WLNEVKCKGNESLMDPCPARWGSHCGHKEDAANCT 1024
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
QY 176 LIPIYW-SNVRGRDDENILLCEKDIMGVCPQKMAAAVTC5 217
RESULT 7
ID 007900 PRELIMINARY; PRT: 1156 AA.
AC 007900;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE M30 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93380506.
RA IAW S.A., MICKLEM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C., MASON D.Y.;
RT "A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily."
RL Eur. J. Immunol. 23:2320-2325(1993).
CP EMBL: 222970; CA80543.1; -
PF PRAM: PF00530; SRCR: 9; -
DR PRINTS: PR00258; SPERACTRCPTR.
KW Antigen; Signal.
FT SIGNAL 1 40
FT CHAIN 41 1156
SQ SEQUENCE 1156 AA; 125352 MW; 287A07A0 CRC32;
Query Match 32.9%; Score 251; DB 4; Length 1156;
Best Local Similarity 39.8%; Pred. No. 2,51e-38;
Matches 41; Conservative 18; Mismatches 40; Indels 4; Gaps 4;
924 IRLQEGTSCSGREIWHGSGMTGVCDDSDMDLDAQVCCOLGCGPALKAFAEF-GQG 982
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
QY 117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
DB 983 TGPI-WLNEVKCKGNESLMDPCPARWGSHCGHKEDAANCT 1024
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
QY 176 LIPIYW-SNVRGRDDENILLCEKDIMGVCPQKMAAAVTC5 217
RESULT 8
ID 092175 PRELIMINARY; PRT: 754 AA.
AC 092175;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-AUG-1999 (TReMBLrel. 11, Last annotation update)
DE LYSYL OXIDASE-RELATED PROTEIN 2 (LOR2 PROTEIN).
GN LOR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J AND 129/SVJ; TISSUE=MUSCLE;
RX MEDLINE: 99126643.
RA JANG W., HUD A., SPILSON S.V., MILLER W., ROE B.A., MEISLER M.H.;
RT "Comparative sequence of human and mouse BAC clones from the mnd2 region of chromosome 2p13."
RL Genome Res. 9:53-61(1999).
DR EMBL: AF053368; AAC83205.1; -
DR EMBL: AF084363; AAC95338.1; -
DR SEQUENCE 754 AA; 83681 MW; 668B44B2 CRC32;
Query Match 32.7%; Score 250; DB 11; Length 754;
Best Local Similarity 40.2%; Pred. No. 4,45e-38;
Matches 41; Conservative 20; Mismatches 39; Indels 2; Gaps 2;
308 VRLKGAGHGEGRVEYVAKATGWTGCDKRDYDAAASYVCELFGRRLSGARRGQGM 367
117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
QY 117 VRLRGKNEFEYEVYASGVWGTCSHWDSDASYICHQLQGGKGA-KQTFPSGLG 175
DB 368 GAHLSEVRCSGOEPLMRCPKNI-TAECSSHSDAGVGN 408
117 IPIYWSNVRGRDDENILLCEKDIMGVCPQKMAAAVTC5 217
QY 176 LIPIYW-SNVRGRDDENILLCEKDIMGVCPQKMAAAVTC5 217
RESULT 9
ID 095178 PRELIMINARY; PRT: 638 AA.
AC 095178;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE LYSYL OXIDASE-LIKE PROTEIN 2.

GN LOXL2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SPLEEN, PLACENTA;
 RX MEDLINE: 99230328.
 RA SAUX C.J., TRONECKER H., BOGIC L., BRYANT-GREENWOOD G.D., BOYD C.D.,
 RA CSISZAR K.;
 RT "The LOXL2 gene encodes a new lysyl oxidase-like protein and is
 RT expressed at high levels in reproductive tissues.";
 RT J. Biol. Chem. 274:12939-12944(1999).
 [2]
 KP SEQUENCE FROM N.A.
 RC TISSUE-SPLEEN, PLACENTA;
 RC JOURDAN-LE SAUX C., CSISZAR K.;
 RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 EMBL: AF117949; AAD3443.1;
 PROSITE: PS00420; SPERACT-RECEPTOR; 1.
 SEQUENCE 638 AA; 71101 MW; C23E6A4 CRC32;
 Query Match 31.8%; Score 243; DB 4; Length 638;
 Best Local Similarity 38.2%; Pred. No. 2,41e-36;
 Matches 39; Conservative 22; Mismatches 38; Indels 3; Gaps 3;
 Db 125 VRLRGAGYIGEGRYEVLKNGEMGTVCDDKMDVLSAVYVRELGFSAKEAVTGSRL-GQG 183
 QY 117 VRLRGKNEFEETVEYASGVGTGCSHMDSDASYICHOLQLG-GKIAKQTFPSGLG 175
 Db 184 IGIHLNEIOCTNEKSIIDCKFNA-ESQGNHEEDAGVRN 224
 QY 176 LIPIYMSNVRCRDEENILLCCKDIWOGVCPQKMAAAVTC 217
 RESULT 10
 ID 029113 PRELIMINARY; PRT; 774 AA.
 AC 09YAKO;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE LYSYL OXIDASE-RELATED PROTEIN.
 GN MS9-14.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SAITO H., PAPACONSTANTINOU J., SATO H., GOLDSTEIN S.;
 RA "Regulation of a Novel Gene Encoding a Lysyl Oxidase-Related Protein
 RA in Cellular Adhesion and Senescence.";
 RA J. Biol. Chem. 0:0-0(1997).
 EMBL: U89942; AAB49697.1;
 DR PROSITE: PS00420; SPERACT-RECEPTOR; 1.
 SQ SEQUENCE 774 AA; 86724 MW; 1A5C47AA CRC32;
 Query Match 31.8%; Score 243; DB 4; Length 774;
 Best Local Similarity 38.2%; Pred. No. 2,41e-36;
 Matches 39; Conservative 22; Mismatches 38; Indels 3; Gaps 3;
 Db 326 VRLRGAGYIGEGRYEVLKNGEMGTVCDDKMDVLSAVYVRELGFSAKEAVTGSRL-GQG 384
 QY 117 VRLRGKNEFEETVEYASGVGTGCSHMDSDASYICHOLQLG-GKIAKQTFPSGLG 175
 Db 385 IGIHLNEIOCTNEKSIIDCKFNA-ESQGNHEEDAGVRN 425
 QY 176 LIPIYMSNVRCRDEENILLCCKDIWOGVCPQKMAAAVTC 217
 RESULT 11
 ID 029112 PRELIMINARY; PRT; 480 AA.
 AC 029112;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
 GN WCI.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SLAB/B;
 RX MEDLINE: 97364683.
 RA KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
 RT "Mechanisms for variability in a member of the scavenger-receptor
 RT cysteine-rich superfamily.";
 RT Immunogenetics 46:276-282(1997).
 RL EMBL: X99335; CAA67709.1;
 DR PRAM: PF00530; SRCR; 3.
 FT NON_TER 1
 SQ SEQUENCE 480 AA; 52368 MW; 25473921 CRC32;
 Query Match 31.7%; Score 242; DB 6; Length 480;
 Best Local Similarity 38.8%; Pred. No. 4,26e-36;
 Matches 40; Conservative 20; Mismatches 39; Indels 4; Gaps 4;
 Db 193 LRLRGSLGSGRYEVLHSSGWTVCDDSKSLAEAYVCOQGLALALDAF-GPG 251
 QY 117 VRLRGKNEFEETVEYASGVGTGCSHMDSDASYICHOLQLG-GKIAKQTFPSGLG 175
 Db 252 SGGI-WLDEVOCGRRESSLMDCAAGPWGSDCKHEEDAGVRCS 293
 QY 176 LIPIYMSNVRCRDEENILLCCKDIWOGVCPQKMAAAVTC 217
 RESULT 12
 ID 029113 PRELIMINARY; PRT; 804 AA.
 AC 029113;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
 DE SCAVENGER-RECEPTOR PROTEIN PRECURSOR.
 GN WCI.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SLAB/B;
 RX MEDLINE: 97364683.
 RA KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
 RT "Mechanisms for variability in a member of the scavenger-receptor
 RT cysteine-rich superfamily.";
 RT Immunogenetics 46:276-282(1997).
 RL EMBL: X99335; CAA67709.1;
 DR PRAM: PF00530; SRCR; 5.
 DR PRINTS: PR00236; SPERACTRCPTR.
 KW SIGNAL.
 FT SIGNAL 1
 FT CHAIN 23 >804 POTENTIAL.
 SQ SEQUENCE 804 AA; 86627 MW; 64125250 CRC32;
 Query Match 31.7%; Score 242; DB 6; Length 804;
 Best Local Similarity 38.8%; Pred. No. 4,26e-36;
 Matches 40; Conservative 19; Mismatches 40; Indels 4; Gaps 4;
 Db 490 LRLRGSGVSGRYEVLHSSGWTVCDDSKSLAEAYVCOQGLALALDAF-GPG 548
 QY 117 VRLRGKNEFEETVEYASGVGTGCSHMDSDASYICHOLQLG-GKIAKQTFPSGLG 175
 Db 549 NGSI-WLDEVOCGRRESSLMDCAAGPWGSDCKHEEDAGVRCS 590
 QY 176 LIPIYMSNVRCRDEENILLCCKDIWOGVCPQKMAAAVTC 217

```
RESULT 13
ID 09Y4V9 PRELIMINARY: PRT: 1785 AA.
AC 09Y4V9;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE DMBT1/6KB.1 PROTEIN PRECURSOR.
GN DMBT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE: 97434209.
MOLENAUER J., WIEMANN S., SCHEURLEN W., KORN B., HAYASHI Y.,
WILGENBUS K.K., VON DEIMLING A., POUSTRKA A.;
EMBL: A7000342; CAA04019.1;
RL "DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-
26.1 is deleted in malignant brain tumours."
Nat. Genet. 17:32-39(1997).
DR EMBL: A7000342; CAA04019.1;
FT SIGNAL.
KW SIGNAL.
SQ SEQUENCE 1785 AA; 193941 MW; 9259828E CRC32;

Query Match 31.4%; Score 240; DB 4; Length 1785;
Best Local Similarity 38.2%; Pred. No. 1.33e-35;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

DB 623 LRLVNGDRCQGVREYLVGSGWGTCDYDNDANVVCQGLCGGAMSPGNARF-GQG 661
QY 117 VRLRGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 682 SCPIVLDVRCSGHESYLVSCPHNGWLSHNGCHHEDAGVICS 723
QY 176 LPIYVSNVRCGDEENILLCERDINQGVCPQKMAAVTCS 217

RESULT 14
ID 09Y211 PRELIMINARY: PRT: 1785 AA.
AC 09Y211;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE DMBT1 PROTEIN.
MOLENAUER J., WIEMANN S., SCHEURLEN W., KORN B., HAYASHI Y.,
WILGENBUS K.K., VON DEIMLING A., POUSTRKA A.;
EMBL: A7000342; CAA04019.1;
RL "DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-
26.1 is deleted in malignant brain tumours."
Nat. Genet. 17:32-39(1997).
DR EMBL: A7000342; CAA04019.1;
FT SIGNAL.
KW SIGNAL.
SQ SEQUENCE 1785 AA; 193941 MW; 9259828E CRC32;

Query Match 31.4%; Score 240; DB 4; Length 1785;
Best Local Similarity 38.2%; Pred. No. 1.33e-35;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

DB 623 LRLVNGDRCQGVREYLVGSGWGTCDYDNDANVVCQGLCGGAMSPGNARF-GQG 661
QY 117 VRLRGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 682 SCPIVLDVRCSGHESYLVSCPHNGWLSHNGCHHEDAGVICS 723
QY 176 LPIYVSNVRCGDEENILLCERDINQGVCPQKMAAVTCS 217

RESULT 15
ID 029110 PRELIMINARY: PRT: 369 AA.
AC 029110;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
GN WCL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SLAB/B;
RX MEDLINE: 97364683.
KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
RT "Mechanisms for variability in a member of the scavenger-receptor
cysteine-rich superfamily."
ImmunoGenetics 46:276-282(1997).
RL EMBL: X99333; CAA67707.1;
DR PROSITE: PS00420; SPBRCT_RECEPTOR. 2.
DR PFM: PF00530; SRCR; 3.
FT NON_TER 1
FT NON_TER 369
SQ SEQUENCE 369 AA; 39710 MW; C669BB4B CRC32;

Query Match 31.3%; Score 239; DB 6; Length 369;
Best Local Similarity 37.9%; Pred. No. 2.34e-35;
Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;
```

```
DR EMBL: AB020829; BAA78577.1; JOINED.
DR EMBL: AB020830; BAA78577.1; JOINED.
DR EMBL: AB020831; BAA78577.1; JOINED.
DR EMBL: AB020832; BAA78577.1; JOINED.
DR EMBL: AB020833; BAA78577.1; JOINED.
DR EMBL: AB020834; BAA78577.1; JOINED.
DR EMBL: AB020835; BAA78577.1; JOINED.
DR EMBL: AB020836; BAA78577.1; JOINED.
DR EMBL: AB020837; BAA78577.1; JOINED.
DR EMBL: AB020838; BAA78577.1; JOINED.
DR EMBL: AB020839; BAA78577.1; JOINED.
DR EMBL: AB020840; BAA78577.1; JOINED.
DR EMBL: AB020841; BAA78577.1; JOINED.
DR EMBL: AB020842; BAA78577.1; JOINED.
DR EMBL: AB020843; BAA78577.1; JOINED.
DR EMBL: AB020844; BAA78577.1; JOINED.
DR EMBL: AB020845; BAA78577.1; JOINED.
DR EMBL: AB020846; BAA78577.1; JOINED.
DR EMBL: AB020847; BAA78577.1; JOINED.
DR EMBL: AB020848; BAA78577.1; JOINED.
DR EMBL: AB020849; BAA78577.1; JOINED.
DR EMBL: AB020850; BAA78577.1; JOINED.
SQ SEQUENCE 1785 AA; 193991 MW; 479E75D8 CRC32;

Query Match 31.4%; Score 240; DB 4; Length 1785;
Best Local Similarity 38.2%; Pred. No. 1.33e-35;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

DB 623 LRLVNGDRCQGVREYLVGSGWGTCDYDNDANVVCQGLCGGAMSPGNARF-GQG 661
QY 117 VRLRGKNEEGVEYVYASGVWGTVCSSHWDDSDASVICHQQLG-GKGIKQTPSSGLG 175
DB 682 SCPIVLDVRCSGHESYLVSCPHNGWLSHNGCHHEDAGVICS 723
QY 176 LPIYVSNVRCGDEENILLCERDINQGVCPQKMAAVTCS 217

RESULT 15
ID 029110 PRELIMINARY: PRT: 369 AA.
AC 029110;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
GN WCL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SLAB/B;
RX MEDLINE: 97364683.
KANAN J.H.C., NAYEM N., BINNS R.M., CHAIN B.M.;
RT "Mechanisms for variability in a member of the scavenger-receptor
cysteine-rich superfamily."
ImmunoGenetics 46:276-282(1997).
RL EMBL: X99333; CAA67707.1;
DR PROSITE: PS00420; SPBRCT_RECEPTOR. 2.
DR PFM: PF00530; SRCR; 3.
FT NON_TER 1
FT NON_TER 369
SQ SEQUENCE 369 AA; 39710 MW; C669BB4B CRC32;

Query Match 31.3%; Score 239; DB 6; Length 369;
Best Local Similarity 37.9%; Pred. No. 2.34e-35;
Matches 39; Conservative 20; Mismatches 40; Indels 4; Gaps 4;
```

Thu Mar 16 07:57:51 2000

US-09-147-947-6-03.rspt

Page 6

OY 176 LPIYW-SNVRCRDEENILCEKDIMQSGVCPQKMAAVTCS 217

Search completed: Mon Mar 13 10:24:48 2000
Job time : 19 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

ch-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:28:46 2000; Maspar time 6.74 Seconds
355.087 Million cell updates/sec
Tabular output not generated.

Title: >US-09-147-947-6
Description: (227-327) from US09147947A-pep (4 of 6)
Perfect Score: 757
Sequence: 1 IRLAGSSVHEGRVELYHAG.....KSWGHRNGHKGADAGVSC 101

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 27.982; Variance 104.818; scale 0.267

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

lt	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	822	1 W9087	Human serine protease	1.56e-71
2	757	100.0	875	1 W8361	Human neutrolypsin.	4.88e-66
3	708	93.5	761	1 W8362	Mouse serine protease	4.88e-66
4	708	93.5	761	1 W9088	Bovine sol. scavenger	9.57e-36
5	435	57.5	453	1 R27036	Scavenger receptor pro	5.59e-35
6	435	57.5	453	1 R0509	Human scavenger recept	5.59e-35
7	428	56.5	451	1 R40802	Human sol. scavenger r	5.59e-35
8	428	56.5	451	1 R27035	Human sol. scavenger r	5.59e-35
9	428	56.5	451	1 W08077	Type I macrophage scav	7.19e-35
10	427	56.4	451	1 R21512	Human macrophage recept	7.19e-35
11	427	56.4	451	1 W19708	Macrophage scavenger r	3.13e-33
12	412	54.4	585	1 R40172	Sequence of an immunor	3.13e-33
13	412	54.4	585	1 R40213	Sequence of a 90k tumo	3.13e-33
14	412	54.4	585	1 R36533	Gp85-97 clone 18 prod	3.13e-33
15	412	54.4	585	1 W81363	Human cytokine regulat	3.13e-33
16	412	54.4	585	1 R41359	Tumour associated 90k	3.13e-33
17	402	53.1	666	1 W64590	Human SRCR protein fra	3.87e-32
18	402	53.1	666	1 W64591	Human SRCR protein fra	3.87e-32
19	373	49.3	574	1 R40784	Sequence of cyclophilli	5.51e-29
20	373	49.3	574	1 W81362	Mouse cytokine regulat	5.51e-29
21	365	48.2	1290	1 W07609	Rat von Ebner's gland	4.06e-28
22	356	47.0	347	1 W68200	Human scavenger recept	3.82e-27
23	356	47.0	347	1 Y13369	Amino acid sequence of	3.82e-27

24	356	47.0	347	1 W64537	Human liver cell clone	3.82e-27
25	355	46.9	489	1 W39749	Mouse macroSR protein.	4.91e-27
26	355	46.9	518	1 W03561	Macrophage receptor wi	4.91e-27
27	338	44.6	495	1 W39747	Human macroSR protein	3.35e-25
28	338	44.6	520	1 W39748	Human macroSR protein	3.35e-25
29	338	44.6	520	1 Y04374	Human macrophage recep	3.35e-25
30	125	16.5	798	1 R57283	Bovine enterokinase.	3.04e-03
31	112	14.8	349	1 W35849	Human CD5 for use in f	4.93e-02
32	88	11.6	356	1 W46917	Amino acid sequence of	6.74e+00
33	83	11.0	652	1 W80610	S. pneumoniae liase.	1.79e-01
34	83	11.0	1167	1 W17700	CryET4.	1.79e-01
35	83	11.0	1167	1 W35258	Bacillus thuringiensis	1.79e-01
36	83	11.0	1167	1 W87632	CryET4 protein sequenc	1.79e-01
37	83	11.0	1167	1 R54073	CryET4.	1.79e-01
38	81	10.7	323	1 R54944	Cuphea hookeriana thlo	2.63e+01
39	80	10.6	1722	1 W38429	Human dendritic cell r	3.18e-01
40	80	10.6	1723	1 W00645	Mouse DEC-205.	3.18e-01
41	80	10.6	1743	1 W00644	Human DEC-205.	3.18e-01
42	79	10.4	1168	1 R89493	CryIF class toxin 91C2	3.85e+01
43	77	10.2	980	1 W69741	SAPAP2 protein.	5.62e+01
44	77	10.2	980	1 W69743	SAPAP1 protein.	5.62e+01
45	77	10.2	1130	1 R13436	Merosin M polypeptide.	5.62e+01

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	W9087	standard; Protein; 822 AA.
AC	W9087	
DT	13-MAY-1999	(first entry)
DE	Human serine protease BSSP-3.	
KW	Serine protease; BSSP-3; brain tissue.	
OS	Homo sapiens.	
PN	W09905290-A1.	
PD	04-FEB-1999.	
PE	24-JUL-1998; J03324.	
PR	24-JUL-1997; JP-213969.	
PA	(SUNR) SUNTORY LTD.	
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;	
DR	WPI; 99-142942/12.	
DR	N-PSDB; X19024.	
PT	New serine protease expressed in brain tissue - used in screening	
PF	for potential serine protease inhibitors for drug use	
PS	Claim 1: Page 61-65; 69pp; Japanese.	
CC	The present sequence is a serine protease designated BSSP-3, which	
CC	is isolated from human brain tissue. Transformants may be used to	
CC	produce the enzyme or its partial sequences. Products from the present	
CC	invention are used for screening for potential peptide or non-peptide	
CC	serine protease inhibitors or expression regulators for use as drugs.	
SO	Sequence 822 AA;	

Query Match	100.0%	Score 757;	DB 1;	Length 822;
Best Local Similarity	100.0%;	Pred. No. 1.56e-71;		
Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Db	227	IRLAGSSVHEGRVELYHAGGWTVCDDDDDAEYICRGLGSLAKAWHOAYFEBS	286	
Qy	227	IRLAGSSVHEGRVELYHAGGWTVCDDDDDAEYICRGLGSLAKAWHOAYFEBS	286	
Db	287	GPVMLDEVRCIGNELSIEQCRKSSWGSHNGHKGADAGVSC 327		
Qy	287	GPVMLDEVRCIGNELSIEQCRKSSWGSHNGHKGADAGVSC 327		
RESULT	2			
ID	W83361	standard; Protein; 875 AA.		
AC	W83361			
DT	17-FEB-1999	(first entry)		
DE	Human neutrolypsin.			
KW	Human; neutrolypsin; tumour inhibition; neurological disease;			
KW	lung disease; gene therapy; drug development; stroke; brain injury;			
KW	neurodegeneration; neuroinflammatory disease; multiple sclerosis;			
KW	epilepsy; hypoxia; ischaemia; nerve transection; neovascularogenesis;			

KW emphysema; bronchitis.
 OS Homo sapiens.
 PN M09849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sondegger P.
 DR MPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 20-24: 50pp; English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 875 AA;

Query Match 100.0%; Score 757; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 1,56e-71;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 280 IRLAGSSVHGRVLYHAGWGTVCDQMDADAEVTCRLGSLGAKAMHQAFFEGS 339
 QY 227 IRLAGSSVHGRVLYHAGWGTVCDQMDADAEVTCRLGSLGAKAMHQAFFEGS 286
 Db 340 GPVMDDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 380
 QY 287 GPVMDDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 327

RESULT 3
 ID W83362 standard; Protein: 761 AA.
 AC W83362;
 DT 17-FEB-1999 (first entry)
 PT Mouse neurotysin.
 KW Mouse; neurotysin; tumour inhibition; neurological disease;
 KW lung disease; gene therapy; drug development; stroke; brain injury;
 KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 KW epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 KW emphysema; bronchitis.

OS Mus musculus.
 PN M09849322-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sondegger P.
 DR MPI: 99-009438/01.
 DR N-PSDB: V72590.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 29-32: 50pp; English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 761 AA;

Query Match 93.5%; Score 708; DB 1; Length 761;
 Best Local Similarity 91.0%; Pred. No. 4.88e-66;
 Matches 91; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 166 IRLVGNNGHGRVLYHAGWGTVCDQMDADADVICRLGSLGAKAMHQAFFEGS 225
 QY 227 IRLAGSSVHGRVLYHAGWGTVCDQMDADAEVTCRLGSLGAKAMHQAFFEGS 286
 Db 226 GPILLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 265
 QY 287 GPVMDDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 326

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088;
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 KW Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN M09905290-A1.
 PD 04-FEB-1999.
 PF 24-JUL-1997; J03324.
 PR (SONR) SUNTORY LTD.
 PA Tsunuoaka N, Yamaguchi N, Yamashiro K;
 PI Tsunuoaka N, Yamaguchi N, Yamashiro K;
 DR MPI: 99-142942/12.
 DR N-PSDB: X19027.

PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1: Page 51-54: 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 761 AA;

Query Match 93.5%; Score 708; DB 1; Length 761;
 Best Local Similarity 91.0%; Pred. No. 4.88e-66;
 Matches 91; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 166 IRLVGNNGHGRVLYHAGWGTVCDQMDADADVICRLGSLGAKAMHQAFFEGS 225
 QY 227 IRLAGSSVHGRVLYHAGWGTVCDQMDADAEVTCRLGSLGAKAMHQAFFEGS 286
 Db 226 GPILLDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 265
 QY 287 GPVMDDEVRCGTGNELSTIEQCPKSSWGEHNCGHKEDAGVSC 326

RESULT 5
 ID R27036 standard; Protein: 453 AA.
 AC R27036;
 DT 16-FEB-1993 (first entry)
 DE Bovine sol. scavenger receptor.
 KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.
 OS Bos taurus.
 FH Key
 FT domain 1. 50

```

FT      domain /note="cytoplasmic"
FT      domain /note="transmembrane"
FT      domain /note="spacer"
FT      domain /note="alpha helical coiled coil"
FT      domain /note="collagen binding"
FT      domain /note="Cys-rich"
FT      domain /note="Cys-rich"
PN      MO9214482-A
PD      03-SEP-1992
PE      21-FEB-1992; 001370.
PR      22-FEB-1991; US-662227.
      (MASI) MASSACHUSETTS INST TECHNOLOGY.
      Krieger M;
      WPI: 92-315935/38.
      N-PSDB: Q28540.
PT      Treatment and diagnosis of endotoxaemia and related disease
PT      states - using a polypeptide fragment of the extracellular
PT      portion of a macrophage scavenger receptor protein
PS      Disclosure: Page 38; 56pp: English.
CC      The bovine scavenger receptor protein sequence was deduced from the
CC      DNA sequence obt'd. by screening a bovine lung cDNA library with
CC      degenerate probes designed based on fragments of the purified
CC      scavenger receptor protein. The scavenger receptor protein has a
CC      binding capacity for acetylated low density lipoprotein (LDL) and
CC      binds endotoxin. Fragments of the extracellular portion of the
CC      scavenger receptor protein inactivate endotoxin-related substances
CC      and are used in the diagnosis and treatment of endotoxaemia. They
CC      may be administered to patients at high risk of symptomatic or
CC      endotoxic shock. The fragments may also be fixed to inert supports
CC      for purification purposes. The fragments may be modified to have
CC      greater binding affinity for the endotoxin-related substance than
CC      the native scavenger receptor protein, or to more effectively
CC      neutralise the toxic or pathogenic effects of mols. that bind
CC      to the scavenger receptor protein or of organisms which express such mols.
CC      See also R27035.
SQ      Sequence 453 AA;

Query Match      57.5%; Score 435; DB 1; Length 453;
Best Local Similarity 54.5%; Pred. No. 9,57e-36;
Matches 55; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

DB      352 VRLVGGSGPHEGVEILFHGQMGTCVDDRMELRGLVCRSLGKGVQSVHKKRAYGKGT 411
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      227 IRLAGSSVHEGVELYHAGQMGTCVDDQMDADAEVLCRLGSLGIAKAMQAYFGEGS 286

412 GPIWLNVEFCFPRESSIEECRIROMGVNRCASHDEDAVGTCT 452
      ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      287 GPVMLDEVRCGTGNELSIEQCPKSSWGHNGKHEDAGVSCT 327

RESULT 6
ID      R05509 standard; protein; 453 AA.
AC      R05509;
DT      23-OCT-1990 (first entry)
DE      Scavenger receptor protein with affinity for acetylated low density
DE      lipoprotein (aLDL).
KW      Acetylated low density lipoprotein; aLDL; atherosclerotic plaque;
KW      ds.
PN      WO9005748-A.
PD      31-MAY-1990.
PE      14-NOV-1989; 005116.
PR      15-NOV-1988; US-272002.
PA      (MASI) Massachusetts Inst Tech.
PI      Engellenner TJ;
DR      WPI: 90-193408/25.
DR      N-PSDB: Q04926.
PT      New receptor protein -
PT      has affinity for acetylated low density lipoprotein and
PT      corresponding antibodies and DNA sequences.

```

```

PS      Disclosure; 79pp; English.
CC      Receptor protein, and fragments and analogues thereof may be
CC      immobilised on a support and used in assay and purification of the
CC      aLDL target. Labelled Abs, raised to the protein may be injected
CC      into the vascular system to detect the presence of atherosclerotic
CC      plaques.
SQ      Sequence 453 AA;

Query Match      57.5%; Score 435; DB 1; Length 453;
Best Local Similarity 54.5%; Pred. No. 9,57e-36;
Matches 55; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

DB      352 VRLVGGSGPHEGVEILFHGQMGTCVDDRMELRGLVCRSLGKGVQSVHKKRAYGKGT 411
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      227 IRLAGSSVHEGVELYHAGQMGTCVDDQMDADAEVLCRLGSLGIAKAMQAYFGEGS 286

DB      412 GPIWLNVEFCFPRESSIEECRIROMGVNRCASHDEDAVGTCT 452
      ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      287 GPVMLDEVRCGTGNELSIEQCPKSSWGHNGKHEDAGVSCT 327

RESULT 7
ID      R40802 standard; protein; 451 AA.
AC      R40802;
DT      23-FEB-1994 (first entry)
DE      Human scavenger receptor I.
KW      Human; scavenger receptor; antihuman scavenger receptor antibody;
KW      AHSRA; mammal; rabbit; antiserum; monoclonal; HAT-resistant hybridoma;
KW      arteriosclerosis.
OS      Homo sapiens.
PN      J05192179-A.
PD      03-AUG-1993.
PE      16-AUG-1991; 229728.
PR      27-AUG-1990; JP-222398.
PA      (CHUS) CHUGAI PHARM CO LTD.
DR      WPI: 93-277488/35.
DR      N-PSDB: Q47730.
PT      Anti-human scavenger receptor antibody - useful for monitoring
PT      progress of arteriosclerosis
PT      Claim 2; Page 5-7; 10pp: Japanese.
CC      The sequences given in R40802-03 represent the human scavenger
CC      receptors I and II. These proteins may be used in the production
CC      of antihuman scavenger receptor antibodies (AHSRA). These peptides
CC      were used to immunise a mammal, pref. a rabbit, and anti-serum was
CC      prepared. A monoclonal antibody was prepared from HAT-resistant
CC      hybridoma. These antibodies may be used to determine the progress
CC      of arteriosclerosis.
SQ      Sequence 451 AA;

Query Match      56.5%; Score 428; DB 1; Length 451;
Best Local Similarity 53.5%; Pred. No. 5,59e-35;
Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

DB      350 VRLVGGSGPHEGVEILFHGQMGTCVDDRMELRGLVCRSLGKGVQSVHKKRAYGKGT 409
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      227 IRLAGSSVHEGVELYHAGQMGTCVDDQMDADAEVLCRLGSLGIAKAMQAYFGEGS 286

DB      410 GPIWLNVEFCFPRESSIEECRIROMGVNRCASHDEDAVGTCT 450
      ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      287 GPVMLDEVRCGTGNELSIEQCPKSSWGHNGKHEDAGVSCT 327

RESULT 8
ID      R27035 standard; protein; 451 AA.
AC      R27035;
DT      16-FEB-1993 (first entry)
DE      Human sol. scavenger receptor.
DE      Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.
OS      Homo sapiens.
FH      Key
FT      domain Location/Qualifiers
FT      domain /note="cytoplasmic"
FT      domain /note="cytoplasmic"

```

FT domain /note="transmembrane"
 FT 77..109
 FT /note="spacer"
 FT 110..271
 FT /note="alpha helical coiled coil"
 FT 272..343
 FT /note="collagen binding"
 FT 341..451
 FT /note="Cys-rich"
 PN W09214482-A.
 PD 03-SEP-1992.
 PF 21-FEB-1992; U01370.
 PR 22-FEB-1991; US-662227.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Krieger M;
 DR WPI; 92-315935/38.
 N-PSDB: Q28539.
 Treatment and diagnosis of endotoxaemia and related disease states - using a polypeptide fragment of the extracellular portion of a macrophage scavenger receptor protein
 Disclosure; Page 33; 56pp; English.
 The human scavenger receptor protein sequence was deduced from the DNA sequence obt'd. by screening a platelet cDNA library with degenerate probes designed based on fragments of the purified scavenger receptor protein. The scavenger receptor protein has a binding capacity for acetylated low density lipoprotein (LDL) and binds endotoxin. Fragments of the extracellular portion of the scavenger receptor protein inactivate endotoxin-related substances and are used in the diagnosis and treatment of endotoxaemia. They may be administered to patients at high risk of symptomatic or endotoxic shock. The fragments may also be fixed to inert supports for purificn. purposes. The fragments may be modified to have greater binding affinity for the endotoxin-related substance than the native scavenger receptor protein, or to more effectively neutralise the toxic or pathogenic effects of mols. that bind to the scavenger receptor protein or of organisms which express such mols. See also R27036.
 CC Sequence 451 AA.
 SQ

Query Match 56.5%; Score 428; DB 1; Length 451;
 Best Local Similarity 53.5%; Pred. No. 5.59e-35;
 Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGPHEGRVELLHSGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 409
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 286

Db 410 GPIWLNEVFCFGRSSIEECKIRQWGTIRACSHSDAGVTCT 450
 QY 287 GPVMDDEVRCGTGNELSIEQCPKSSWGEHNGCHKEDAGVSC 327

RESULT 9
 ID W08077 standard; Protein: 451 AA.
 AC W08077.

DE 25-FEB-1997 (first entry)
 DE Type I macrophage scavenger receptor.
 KW Type I macrophage scavenger receptor; SRG; kidney; 293 cell;
 KW cell substrate; cell attachment; atherosclerosis.
 OS Homo sapiens.
 PN W09638725-A1.
 PD 05-DEC-1996.
 PF 30-MAY-1996; U08081.
 PR 30-MAY-1995; US-453117.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Brawner ME; Elshourbagy NAE, Lysko PG;
 DR N-PSDB: 143259.
 DR WPI: 97-034505/03.
 PT New modified human embryonic kidney 293 cells - transfected with a mammalian scavenger receptor gene to enhance ability to attach to a solid support
 PT Disclosure: Page 10-13; 32pp; English.
 PS Human type I (W08077) and type II (W08078) macrophage scavenger

CC receptors can be expressed in human embryonic kidney 293 cells
 CC following transfection of the cells with vectors carrying the
 CC respective scavenger receptor genes (T43259 and T43260). The
 CC transfected cells show an enhanced ability to attach to a solid
 CC support and are useful for screening cpds. for biological activity,
 CC or for identifying antagonists of the scavenger receptor gene, e.g.
 CC to develop agents for treatment of atherosclerosis.
 SQ Sequence 451 AA.

Query Match 56.5%; Score 428; DB 1; Length 451;
 Best Local Similarity 53.5%; Pred. No. 5.59e-35;
 Matches 54; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGPHEGRVELLHSGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 409
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 286

Db 410 GPIWLNEVFCFGRSSIEECKIRQWGTIRACSHSDAGVTCT 450
 QY 287 GPVMDDEVRCGTGNELSIEQCPKSSWGEHNGCHKEDAGVSC 327

RESULT 10
 ID R21512 standard; Protein: 451 AA.
 AC R21512.

DE 19-MAY-1992 (first entry)
 DE Human scavenger receptor type I.
 KW Lipoproteins.
 OS Homo sapiens.
 PN J03290184-A.
 PD 19-DEC-1991.
 PF 06-APR-1990; 090274.
 PR 06-APR-1990; JP-090274.
 PA (CHUS) CHUGAI PHARMACEUTICAL KK.
 DR WPI: 92-051436/07.
 DR N-PSDB: Q21549.

PT New scavenger receptor-producing animal cells - which have been
 PT transformed with vector contg. gene for coding human scavenger
 PT receptor I or II type under control of promoter, etc.
 PS Disclosure; Fig 1; 9pp; Japanese.
 CC The gene encoding the receptor can be expressed from a cytomegalo-
 CC virus promoter in a host cell e.g. CHO cell. The resulting
 CC recombinant scavenger receptor can be used to detect modified lipo-
 CC proteins or modified substances in the blood.
 CC See also R21513.
 SQ Sequence 451 AA.

Query Match 56.4%; Score 427; DB 1; Length 451;
 Best Local Similarity 52.5%; Pred. No. 7.19e-35;
 Matches 53; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Db 350 VRLVGGSGPHEGRVELLHSGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 409
 QY 227 IRLAGSSVHEGRVELYHAGQWGTICDDREVRVGOVVCRSIGYGVQAVHKAHFGGCT 286

Db 410 GPIWLNEVFCFGRSSIEECKIRQWGTIRACSHSDAGVTCT 450
 QY 287 GPVMDDEVRCGTGNELSIEQCPKSSWGEHNGCHKEDAGVSC 327

RESULT 11
 ID W19708 standard; Protein: 451 AA.
 AC W19708.

DE 19-AUG-1997 (first entry)
 DE Macrophage scavenger receptor protein.
 KW LTA; Gram-positive bacteria; cell wall; atherosclerosis; host defence;
 KW septicemia; inhibitor; complement activation; cytokine release; therapy;
 KW nitric oxide production; bacterial infection; septic shock.
 OS Homo sapiens.
 PN US5624904-A.
 PD 29-APR-1997.
 PF 17-NOV-1993; 154365.

PR 17-NOV-1993: US-154365.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI (UYIA) UNIV YALE.
 PI Joiner KA, Krieger M;
 DR WPI: 97-258236/23.
 N-PSDB: T68796.
 PT Treatment of septicemia caused by Gram-positive bacteria - by
 administration of macrophage scavenger receptor protein
 PS Disclosure: Column 23-26: 16pp; English.
 CC This sequence represents the human macrophage scavenger receptor protein
 (MSRP). MSRP binds specifically to the lipoteichoic acid (LTA) residues
 on the Gram-positive bacterial cell wall. MSRP exhibit unusually broad
 binding specificity for polyanionic ligands, and have been implicated in
 atherosclerosis and a variety of host defence functions. This protein can
 be used in the method of the invention. The method of the invention is
 for the treatment of septicemia caused by Gram-positive bacteria. The
 method comprises administration of a MSRP in an amount sufficient to
 inhibit complement activation, cytokine release or nitric oxide
 production induced by LTA released by the Gram-positive bacteria. In
 addition to treating Gram-positive septic shock, the MSRP can be used to
 screen for other compounds for treating Gram-positive septic shock. MSRP
 can also be used to purify, label or detect LTA or LTA-containing cells.
 CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
 inhibit binding of Gram-positive bacteria to MSRP can be used to treat
 CC pathologies such as septicemia, Gram-positive bacterial infection,
 CC Gram-positive septicemia or Gram-positive or Gram-negative septic shock.
 SQ Sequence 451 AA;

Query Match 56.4%; Score 427; DB 1; Length 451;
 Best Local Similarity 53.5%; Pred. No. 7,19e-35;
 Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 350 VRLVGGSGPHEGRVYELHSGQWGTICDDNNEVRGVYVCHSLGPGVQAVHKAHFGCGT 409
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGGSVHEGRVELYHAGQWGTICDDNDAEYICQGLSGTAKMHWQAYFEEGS 286
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 410 GPIMLNEVFCFGRSSIEECKIRQWGTACSHSDACVCT 450
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 287 GPVWLDEVCTGNELSTIECPKSSWGEHNGHKEDAGVCT 327

RESULT 12
 ID R40172 standard; Protein: 585 AA.
 AC R40172;
 DT 08-FEB-1994 (first entry)
 DE Sequence of an immunoregulin (IR)-95 polypeptide.
 PI Tumour associated antigen; breast cancer; cell line CG-5;
 PI ovarian cancer; immunoregulin-95; IR-95.
 PI Homo sapiens.
 Key Location/Qualifiers
 FT peptide 1..17
 FT region 24..125
 FT modified_site 71
 FT /label= glycosylation site
 FT /note="see also AAs 125,192,362,398,551,580"
 PN MO9316180-A.
 PD 19-AUG-1993.
 PF 17-FEB-1993: E00382.
 PR 17-FEB-1992: IT-RM0100.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (UYCH-) UNIV CHIETI ANNUNIO G D.
 PI Azam M, Iacobelli S, Natoli C, Sures I, Ullrich A;
 DR WPI: 93-272884/34.
 N-PSDB: Q46888.
 PT Recombinant DNA - encoding tumour associated antigen,
 PT Immunoregulin-95
 PS Disclosure: Fig 1: 6pp; English.
 CC Immunoregulin-95 is a 90k tumour-associated antigen purified from
 CC the culture fluid of the human breast cancer cell line, CG-5, the
 CC serum of a breast cancer patient, or the ascitic fluid from an
 CC ovarian cancer patient. The native antigen, which has a mol. wt. of

CC 95 kD, is present as a high mol. wt. complex. Homology in the region
 CC of AAs 35-80 of the 90k antigen is found with type I macrophage
 CC scavenger receptor; sea urchin speract receptor; and human
 CC lymphocyte glycoprotein 11/Leu-1.
 SQ Sequence 585 AA;

Query Match 54.4%; Score 412; DB 1; Length 585;
 Best Local Similarity 53.5%; Pred. No. 3,13e-33;
 Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 24 MRLADGATNOGRVEIRYRGOWGTICDNLMDLDASVCRALGFENATOLGRAFGGS 83
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGGSVHEGRVELYHAGQWGTICDDNDAEYICQGLSGTAKMHWQAYFEEGS 286
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 84 GPIMLDEVCTGTEASLADCKSLGWLKSNCRHBRDAGVCT 124
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 287 GPVWLDEVCTGNELSTIECPKSSWGEHNGHKEDAGVCT 327

RESULT 13
 ID R40213 standard; Protein: 585 AA.
 AC R40213;
 DT 03-FEB-1994 (first entry)
 DE Sequence of a 90k tumour-associated antigen, IR-95, capable
 DE of binding to SP-2 (Accession No. I-1083).
 KW Tumour-associated antigen; breast cancer cell line CG-5;
 KW 90 K antigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 1..18
 FT region 24..125
 FT binding_site 69
 FT /label= potential glycosylation site
 FT /note="see also AAs 125,192,363,398,551,580"
 PN MO9316181-A.
 PD 19-AUG-1993.
 PF 17-FEB-1993: E00385.
 PR 17-FEB-1992: IT-RM0099.
 PA (UYCH-) UNIV CHIETI ANNUNIO G D.
 PI (UYNY) UNIV NEW YORK STATE.
 PI Iacobelli S, Natoli C, Schlessinger J;
 DR WPI: 93-272885/34.
 N-PSDB: Q48170.
 DT New 90k tumour-associated antigen, IR-95 - is for use in
 PT diagnosis and therapy of cancer, HIV and auto-immune diseases
 PS Claim 2: pages 45-48; 68pp; English.
 CC 90K antigen has an apparent molecular weight of approx. 95 kD. It is
 CC a tumour associated antigen and is elevated in the serum of patients
 CC with cancer and also in patients with HIV. It reacts with Mab SP-2
 CC which was prod. by immunising mice with proteins that had been
 CC released into tissue culture fluid by human MCF-7 breast cancer
 CC cells. Mab SP-2 cell line is deposited at the Institut Pasteur,
 CC Paris. Accession number 1-1083. 90K is also present in normal
 CC subjects. It is purified from the culture fluid of the human breast
 CC cancer cell line, CG-5. N-terminal sequencing of the 90K antigen
 CC was used to design a 'guesser' nucleotide sequence (Q48171) as a
 CC probe to screen a lambda-gt10 library prepd. from MCF7 polyA+ RNA.
 CC The complete nucleotide sequence of isolated clones is given in
 CC Q48170.
 SQ Sequence 585 AA;

Query Match 54.4%; Score 412; DB 1; Length 585;
 Best Local Similarity 53.5%; Pred. No. 3,13e-33;
 Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

Db 24 MRLADGATNOGRVEIRYRGOWGTICDNLMDLDASVCRALGFENATOLGRAFGGS 83
 :|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 227 IRLAGGSVHEGRVELYHAGQWGTICDDNDAEYICQGLSGTAKMHWQAYFEEGS 286
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 84 GPIMLDEVCTGTEASLADCKSLGWLKSNCRHBRDAGVCT 124
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 287 GPVMDDEVCTGNELSTIEQCPKSSWGEHNCGRKEDAGVCT 327

RESULT 14

ID R36533 standard; Protein; 585 AA.

AC R36533:

DT 19-AUG-1993 (first entry)

DE GP85-97 clone 18 prod.

KW Glycoprotein; lectin; Mac-2; PHA; lymphocytes; leucocagglutinating;

OS phytohemagglutinin; cancer; amplification; PCR.

HM Homo sapiens.

FT Key Location/Qualifiers

FT peptide 1..18

FT protein /note="leader sequence"

FT 19..585

PN WO9308215-A.

29-APR-1993.

15-OCT-1992; U08878.

16-OCT-1991; US-777121.

15-OCT-1992; US-961404.

(CETU) CETUS ONCOLOGY CORP.

PI Caspiti CL, Halsebeck R, Koths KE, Taylor EM, Wang AM;

PT New glyco:protein complex binding to human lectin Mac-2 - also

PT interferes with PHA activation of lymphocytes for treating and

PS preventing cancer; infectious diseases; etc.

CC Claim 1. Page 41; 57pp; English.

CC SK-BR-3 gp97 recovered in partially proteolysed form was denatured

CC and reduced and the 97 and 70 kd mols. were purified using size

CC exclusion HPLC in 0.1 percent SDS. The 97 and 70 kd mols. were

CC digested with Lys-C protease and the resulting peptides purified and

CC sequenced. The N-terminal sequence of gp97 was used to design

CC degenerate primers for use in a PCR reaction with SK-BR-3 mRNA. PCR

CC reactions using primer 2 in combination with primers 1 and 3 gave prods.

CC of 97 and 121 bp. Further PCR using primers 2, 3 and 4 yielded DNA

CC sequences of 740 and 765 respectively. A DNA sequence was obtd. from

CC the PCR prods. Two additional oligonucleotide sequences were

CC synthesised based on this sequence and used to probe a TBP-1 CDNA

CC library to obtain the full-length cDNA sequence that encodes SK-BR-3

CC gp97. Both strands of clone 218 were sequenced. It encodes a novel

CC glycoprotein binds the Mac-2 lectin and interferes with PHA activation

CC of lymphocytes. It can be used for treating or preventing diseases that

CC result from binding of a disease-causing agent to the cell surface of a

CC target cell. The gp. can be used in treatment of cancers, partic.

CC breast cancer. See also R36532.

CC Sequence 585 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 54.4%: Score 412; DB 1; Length 585;
Best Local Similarity 53.5%: Pred. No. 3.13e-33;
Matches 54; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

DB 24 MRLADGATNOGVEIFRGOMGTVCNMDLTDASVCRALGFENATLGRAAFGGS 83
QY 227 IRLAGSSVHEGRVELYHAGOMGTVCDDMDADAEVYICROLGSLGAKAMHQAIFGEGS 286

DB 84 GPIMLDEVCTGTASLADCKSLGWLKSNCRHERDAGVCT 124
QY 287 GPVMDDEVCTGNELSTIEQCPKSSWGEHNCGRKEDAGVCT 327

Search completed: Mon Mar 13 10:28:55 2000
Job time : 9 secs.

FR Key Location/Qualifiers
FT Misc.difference 257 /note="encoded by Cgr"
PN WO9856819-A1.
FT 17-DEC-1998.
PD 12-JUN-1998: U12345.
PR 13-JUN-1997; US-049578.
PI (STRD) UNIV LEIDAN STANFORD JUNIOR.
DR Conboy IB, Jones PP;
DR N-PSDB: V68491.
PT Reduction of pro-inflammatory cytokines in a T cell population -
PT using cytokine regulatory factor (CYTRF), useful in the diagnosis
PT and treatment of pro-inflammatory conditions
PS Disclosure: Page 47-51; 77pp; English.
CC This is the amino acid sequence of human cytokine regulatory factor
CC (CYTRF), also known as cyclophilin C-associated protein (CYCAP).
CC Methods are provided for the modulation of cytokine production by T
CC helper cells using CYTRF (especially human CYTRF). A pharmaceutical
CC composition comprising CYTRF as an active agent is administered in
CC vitro or in vivo, and can act on mature, committed Th1 type T cells
CC to decrease the production of pro-inflammatory cytokines, or to skew
CC the commitment of precursor T helper cells to Th1 or a Th1 cytokine
CC profile. The administration of CYTRF is useful in the diagnosis and
CC treatment of autoimmune diseases (e.g. multiple sclerosis,
CC rheumatoid arthritis, and degenerative joint diseases including
CC ankylosing spondylitis and reactive arthritis), and inflammation
CC caused by bacterial, viral or parasitic infection (e.g. meningitis,
CC herpes encephalitis, viral meningoencephalitis and viral hepatitis),
CC including response to vaccination (e.g. vaccines containing rabies,
CC varicella zoster, and measles) and skin sensitivity (e.g. graft
CC rejection and graft vs host disease). CYTRF signalling and related
CC pathways are also useful for modelling and screening new
CC pharmacological agents. The administration of agents that block
CC CYTRF action (e.g. antibodies) are useful in reducing unwanted
CC allergic responses, especially in asthma.
SQ Sequence 585 AA;

(TM)

Distribution rights by Oxford Molecular Ltd

1 VRLRGKNEFEGETVEVYASG.....KDIWQGGVC P Q K M A A V T C S 101

PAM 150

188963 seqs, 23686106 residues

Listing first 45 summaries

1:geneseqf

Mean 28.207; Variance 100.840; scale 0.280

and is derived by analysis of the total score distribution.

SUMMARIES

t	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	764	100.0	822	1	W95087	Human serine protease	8.43e-76
2	764	100.0	875	1	W83361	Human neurolypsin.	8.43e-76
3	323	42.3	761	1	W83362	Mouse neurileptysin.	8.75e-25
4	323	42.3	761	1	W93088	Mouse serine protease	8.75e-25
5	240	31.4	1785	1	W64591	Human SSCR protein.	1.4e-15
6	239	31.3	489	1	W39749	Mouse macroSR protein.	1.81e-15
7	239	31.3	518	1	W03561	Macrophage receptor v1	1.81e-15
8	237	31.0	451	1	R40802	Human scavenger recept	2.99e-15
9	237	31.0	451	1	R27033	Human sol. scavanger r	2.99e-15
10	237	31.0	451	1	W19708	Macrophage scavenger r	2.99e-15
11	237	31.0	451	1	W08077	Type I macrophage scav	2.99e-15
12	236	30.9	451	1	R21512	Human scavenger recept	3.84e-15
13	229	30.0	666	1	W64550	Human SSCR protein fra	2.22e-14
14	225	29.5	347	1	W68200	Human scavenger recept	6.04e-14
15	225	29.5	347	1	Y13369	Amino acid sequence of	6.04e-14
16	225	29.5	347	1	W64537	Human liver cell clone	7.75e-14
17	224	29.3	453	1	R27036	Scavenger receptor	7.75e-14
18	224	29.3	453	1	R05500	Gp85-97 clone 18 prod.	2.69e-13
19	219	28.7	585	1	R36533	Human cytokine regulat	2.69e-13
20	219	28.7	585	1	W81363	Sequence of an Immunor	2.69e-13
21	219	28.7	585	1	R40172	Tumour associated 90K	2.69e-13
22	219	28.7	585	1	R41359	Sequence of a 90K tumo	2.69e-13
23	219	28.7	585	1	R40213		

[illegible]

ALIGNMENTS

RESULT	1
ID	W99087 standard; Protein; 822 AA

DT 13-MAY-1999 (first entry)

KW Serine protease; BSSP-3; brain tissue.

PN W09905290-A1

PF 24-JUL-1998; J03324.

PA (SUNR) SUNTORY LTD.

DR WPI: 99-142942/12.

New serine protease

PS Claim 1: Page 61-65: 69nm: Japanese
F1 101 potential 3elme p10cease 1111d1c10rs 101 111g use

CC The present sequence is a serine protease designated BSSP-3, which
CC is isolated from human brain tissue. Transformants may be used to

produce the enzyme or its partial sequences. Products from the present invention are used for controlling for rotavirus or non-rotavirus

serine protease inhibitors or expression regulators for use as drugs.

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

Best Local Similarity 100.0%; Pred. No. 8.43e-76;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 2

W83361:
ID scandau, floell,
AC

DE Human neutrotrofin

Human; neurolypsin; tumour inhibition; neurological disease; lung disease; drug development; stroke; brain

KW neurodegeneration; neuroinflammation; multiple sclerosis; neuroprotection

KW emphysema; bronchitis.
 OS Homo sapiens.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1; Page 20-24; 50pp; English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SO Sequence 875 AA:

Query Match 100.0%; Score 764; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 8.43e-76;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 170 VRLRGKNEFEETVEYASGVGWTSSHWDDSDAVICHOLOGKGIKAKQPFSGLG 229
 QY 117 VRLRGKNEFEETVEYASGVGWTSSHWDDSDAVICHOLOGKGIKAKQPFSGLG 176
 Db 230 IPIYWSNRCRGDEENILCEKIDIMOGVCPOKMAAAYTC 270
 QY 177 IPIYWSNRCRGDEENILCEKIDIMOGVCPOKMAAAYTC 217

RESULT 3
 ID W83362 standard; Protein; 761 AA.
 AC W83362;
 PT 17-FEB-1999 (first entry)
 PI Mouse neurotysin.

KW Mouse; neurotysin; tumour inhibition; neurological disease;
 KW lung disease; gene therapy; drug development; stroke; brain injury;
 KW neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 KW epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 KW emphysema; bronchitis.

OS Mus musculus.
 PN W09849322-A1.
 PD 05-NOV-1998.
 PE 24-APR-1998; IB0625.
 PR 26-APR-1997; CH-000966.
 PA (SOND/) SONDEREGGER P.
 PI Sonderegger P.
 DR WPI: 99-009438/01.
 DR N-PSDB: V72590.

PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1; Page 29-32; 50pp; English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases, e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SO Sequence 761 AA:

Query Match 42.3%; Score 323; DB 1; Length 761;
 Best Local Similarity 45.5%; Pred. No. 8.75e-25;
 Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Db 166 IRLVGNSGHEGRVELYHAGWGITCDQMDNADADVICROLGSLGAKAHQAHP-GE 224
 QY 117 VRLRGKNEFEETVEYASGVGWTSSHWDDSDAVICHOLOGKGIKAKQPFSGLG 175
 Db 225 SGPIILDEVRCGTNETSTPOCKSSWGEHNCGRKEDAGVSC 265
 QY 176 IPIYWSNRCRGDEENILCEKIDIMOGVCPOKMAAAYTC 216

RESULT 4
 ID W99088 standard; Protein; 761 AA.
 AC W99088;
 DT 13-MAY-1999 (first entry)
 DE Mouse serine protease BSSP-3.
 KW Serine protease; BSSP-3; Brain tissue.
 OS Mus sp.
 PN W09905290-A1.
 PD 04-FEB-1999.
 PE 24-JUL-1998; J03324.
 PR 24-JUL-1997; JP-213969.
 PA (SDNR) SONTORY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 DR N-PSDB: X19027.

PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1; Page 51-54; 69pp; Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transformants may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SO Sequence 761 AA:

Query Match 42.3%; Score 323; DB 1; Length 761;
 Best Local Similarity 45.5%; Pred. No. 8.75e-25;
 Matches 46; Conservative 19; Mismatches 34; Indels 2; Gaps 2;

Db 166 IRLVGNSGHEGRVELYHAGWGITCDQMDNADADVICROLGSLGAKAHQAHP-GE 224
 QY 117 VRLRGKNEFEETVEYASGVGWTSSHWDDSDAVICHOLOGKGIKAKQPFSGLG 175
 Db 225 SGPIILDEVRCGTNETSTPOCKSSWGEHNCGRKEDAGVSC 265
 QY 176 IPIYWSNRCRGDEENILCEKIDIMOGVCPOKMAAAYTC 216

RESULT 5
 ID W64591 standard; Protein; 1785 AA.
 AC W64591;
 DT 23-OCT-1998 (first entry)
 DE Human SRCR protein.
 KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
 KW nervous system; medullo-blastoma; glioma; breast; detection;
 KW autoantibody; ss.
 OS Homo sapiens.

PN MO9830687-AZ.
PD 16-JUL-1998.
PL US-JAN-1988; D00096.
PR 08-JUL-1997; DE-03097.
PS 09-JAN-1997; DE-000519.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Mollenhauer J, Poustka A:
DR WPI: 98-399136/34.
NR N-PDSB: V49652.
PT Proteins containing scavenger receptor, cysteine rich domain -
PF useful for diagnosis and treatment of tumours
PS Claim 2; Fig 2; 54pp; German.
CC This sequence represents a human protein which contains a SNCR (scavenger
receptor, cysteine-rich) domain. The gene and encoded protein can be used
to diagnose or treat tumours, particularly of the nervous system
(medullo-blastoma or glioma) or breast. The DNA sequence and probes
derived from it, are used to identify genes that express SNCR-domain
containing proteins, to determine the form in which these proteins exist
and to assess the significance of individual forms on cellular
properties. The protein can be used to detect the presence of
autoantibodies and antibodies which regulate its expression.

CC Sequence 1785 AA.

Query Match 31.4%; Score 240; DB 1; Length 1785;
Best Local Similarity 38.2%; Pred. No.1.41e-15;
Matches 39; Conservative 18; Mismatches 43; Indels 2; Gaps 2;

Db 623 LRLVNGRCRGQREVLYLRSWGTVCDYDWDNDANWVCROLGCGMAMPARNF-GQG 681
::| | : : : | : | | | | | | | | : : : : | | |
QY 117 VLRLGKGRNEFEGETVEYVASVGWTGVSCHMDSDASVICHQLGLE-KGIKIQKTFPSLG 175
::| | : | | | | : : : : : : : : : : : : | | |

Db 682 SGPIVDLVACSGHESITLMSCPHNGLSHNCGHEDAGYTCS 723
|| : || | | | : : : : : : : : : : : : | | |

QY 176 LIPIYMSNVRCRDENILCERDDIMQGGVCPOKMAAAVTCS 217

RESULT 6
ID W39749 standard; Protein: 489 AA.
AC W39749;
DT 20-MAY-1998 (first entry)
DE Mouse macroSR protein.
KW Marco scavenger receptor; MacroSR; mouse; antibodies; treatment;
diagnosis; cardiovascular disease; macrophage; bacterial infection;
immune cell related host cell disorder; pancreatitis; organ failure;
endotoxaemia; agonist; antagonist; dysfunction.
OS Mus sp.
EP-808899-AA;
26-NOV-1997
19-MAY-1997; 303394.
04-FEB-1997; US-794795.
23-MAY-1996; US-017699.
(SMIK) SMITHKLIN BEECHAM CORP.
PI Adamou JE, Elshourbagy NAM, Gross MS, Lyoko PG;
WPI: 98-001786/01.
PT Human Macro scavenger receptor - used to develop products for
treating e.g. cardiovascular disorders, septic shock, bacterial
infections etc.
PT Disclosure: Fig 1: 45pp; English.
PS This sequence represents a mouse Macro scavenger receptor (MacroSR)
which is used to confirm the presence of a novel human macro scavenger
receptor (HmacroSR). This polypeptide and resulting antibodies can be
used to treat cardiovascular diseases, including atherosclerosis,
hypertension myocardial and cerebral infarction, angina, organ failure,
stroke, gangrene, and loss of function in the extremities. They can also
be used to treat or diagnose various macrophage and other immune cell
related host defense disorders, septic shock, pancreatitis, multiple
organ failure, endotoxaemia and infections caused by gram negative and
gram positive bacteria. Agonists and antagonists may also be used to
treat these diseases. The polynucleotide can be used to detect mutated
forms of HmacroSR which may be associated with a dysfunction and can
define a diagnosis of a disease (or susceptibility to a disease)
resulting from altered HmacroSR expression.

Sequence 489 AA;

```

Db      394 VRIMGTR--GRAEYYNNEMGTICDDMDNDATFVCRLTGYG-RCRA-LSSYG-GS 448
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      117 VRLRGKNEFEGETVEYVAVSGWGVTCSSHMDSDASVICHQLDGGKGIAKOTPFSLGL 176
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      449 GNI-WLDNVNCRGTENSLMCSKRWGNHNCVHNEDGEGVCS 489
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      177 IPIYW-SNVRCRGEENTLLCEKDIMGVGCPOKMAAAVTCS 217
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
ID      W03561 standard; Protein; 518 AA.
DC      W03561.
DT      22-JAN-1997 (first entry)
DE      Macrophage receptor with collagenous domain (MARCO).
KW      Macrophage receptor-collagenous; MARCO; immune defense;
        phagocytosis.
OS      Mus sp.
FH      Key
FT      Location/Qualifiers
FT      domain
           1..49
           /label= Domain-I
           /note="hydrophilic domain not containing a signal
           peptide"
FT      domain
           50..74
           /label= Domain-II
           /note="hydrophobic domain"
FT      domain
           75..149
           /label= Domain-III
           /note="hydrophilic domain"
FT      modified_site
           87..89
           /label= glycosylation
           /note="putative N-glycosylation site"
FT      modified_site
           138..140
           /label= glycosylation
           /note="putative glycosylation site"
FT      domain
           150..419
           /label= Domain-IV
           /note="collagenous structure characterised by 89
           Gly-X-X repeats with one interruption
           (Ala-Glu-Lys at 174-176)"
FT      domain
           420..518
           /label= Domain-V
           /note="globular C-terminal domain"
FT      FT
PN      W09626219-A1.
PD      29-AUG-1996.
PE      19-FEB-1996; F10091.
PR      21-FEB-1995; US-392367.
PA      (ELOM/) ELONMAA O.
         (KANG/) KANGAS M.
         (TRYG/) TRYGVASON K.
         Elomaa O, Kangas M, Trygvason K;
         WPI: 96-402322/40.
DR      N-PDB: T39340.
DI      New macrophage receptor, MARCO, that binds bacteria but not yeast -
PT      and related DNA, anti-sense probes and antibodies, involved in
PT      immune defence and phagocytosis
PS      Claim 10; Fig 2A-B; 62pp; English.
CC      A novel macrophage receptor with a collagenous domain, named MARCO
CC      (W03561), is a membrane-bound trimeric protein which is expressed
CC      strongly after birth in a subset of macrophages in mouse spleen and
CC      lymph nodes. It binds to bacteria but not to yeast, and probably
CC      plays a role in immune defence and/or phagocytosis. The amino acid
CC      sequence of MARCO was deduced from DNA (see also t39340) obid. from
CC      a mouse macrophage cDNA library. Recombinant MARCO can be produced
CC      in transformed host cells. It is used to raise antibodies useful
CC      in detecting MARCO or interfering with its function.
SQ      Sequence 518 AA:

Query Match       31.3%; Score 239; DB 1; Length 489;
Best Local Similarity 39.2%; Pred. No. 1,81e-15;
Matches 40; Conservative 21; Mismatches 34; Indels 7; Gaps 5;

```


CC inhibit complement activation, cytokine release or nitric oxide
 CC production induced by LTA released by the Gram-positive bacteria. In
 CC addition to treating Gram-positive septic shock, the MSRP can be used to
 CC screen for other compounds for treating Gram-positive septic shock. MSRP
 CC can also be used to purify, label or detect LTA or LTA-containing cells.
 CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
 CC inhibit binding of Gram-positive bacteria to MSRP can be used to treat
 CC pathologies such as septicemia, Gram-positive bacterial infection,
 CC Gram-positive septicemia or Gram-positive or Gram-negative septic shock.
 SQ Sequence 451 AA;

Query Match 31.0%; Score 237; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 2,99e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

350 VRLVGGSGPHEGRVEILHSGOMGTICDDREVEVGVGVRSLGYVQVAKAAHF-GQG 408
 117 VRLRGKNEFEIGVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 409 TGPI-WLNEVFCFGRSSIEECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILLCERKIDWOGVCPQKMAAAVTCS 217

RESULT 11
 ID W08077 standard; Protein: 451 AA.

AC W08077;
 DT 25-FEB-1997 (first entry)
 DE Type I macrophage scavenger receptor.
 KM Type I macrophage scavenger receptor; SRG; kidney; 293 cell;
 OS Homo sapiens; cell attachment; atherosclerosis.
 PN W09638725-A1.
 PD 05-DEC-1996.
 PR 30-MAY-1995; US-453117.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Brainer ME, Elshourbagy NAE, Lysko PG;
 DR N-PSDB; T43259.
 PT New modified human embryonic kidney 293 cells - transfected with a
 PT mammalian scavenger receptor gene to enhance ability to attach to a
 PT solid support
 PS Disclosure: Page 10-13; 32pp; English.
 CC Human type I (W08077) and type II (W08078) macrophage scavenger
 CC receptors can be expressed in human embryonic kidney 293 cells
 CC following transfection of the cells with vectors carrying the
 CC respective scavenger receptor genes (T43259 and T43260). The
 CC transfected cells show an enhanced ability to attach to a solid
 CC support and are useful for screening cpds. for biological activity,
 CC or for identifying antagonists of the scavenger receptor gene, e.g.
 CC to develop agents for treatment of atherosclerosis.
 SQ Sequence 451 AA;

Query Match 31.0%; Score 237; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 2,99e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHEGRVEILHSGOMGTICDDREVEVGVGVRSLGYVQVAKAAHF-GQG 408
 117 VRLRGKNEFEIGVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 DB 409 TGPI-WLNEVFCFGRSSIEECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILLCERKIDWOGVCPQKMAAAVTCS 217

RESULT 12
 ID R21512 standard; Protein: 451 AA.
 AC R21512;
 DT 19-MAY-1992 (first entry)
 DE Human scavenger receptor type I.
 KW Lipoproteins.

OS Homo sapiens.
 PN J03290184-A.
 PD 19-DEC-1991.
 PF 06-APR-1990; 090274.
 PR 06-APR-1990; JP-090274.
 PA (CHUS) CHUGAI PHARMACEUTICAL KK.
 DR WPI: 92-051436/07.
 DR N-PSDB; Q21549.
 PT New scavenger receptor-producing animal cells - which have been
 PT transformed with vector contg. gene for coding human scavenger
 PT receptor I or II type under control of promoter, etc.
 PS Disclosure: Fig 1; 3pp; Japanese.
 CC The gene encoding the receptor can be expressed from a cytomagalo-
 CC virus promoter in a host cell e.g. CHO cell. The resulting
 CC recombinant scavenger receptor can be used to detect modified lipo-
 CC proteins or modified substances in the blood.
 CC See also R21513.
 SQ Sequence 451 AA;

Query Match 30.9%; Score 236; DB 1; Length 451;
 Best Local Similarity 37.9%; Pred. No. 3,84e-15;
 Matches 39; Conservative 17; Mismatches 43; Indels 4; Gaps 4;

DB 350 VRLVGGSGPHEGRVEILHSGOMGTICDDREVEVGVGVRSLGYVQVAKAAHF-GQG 408
 117 VRLRGKNEFEIGVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175
 DB 409 TGPI-WLNEVFCFGRSSIEECKIRMGTRACSHSDAGVTCT 450
 176 LPIIYW-SNVRCRGDEENILLCERKIDWOGVCPQKMAAAVTCS 217

RESULT 13

ID W64590 standard; Protein: 666 AA.
 AC W64590;
 DT 23-OCT-1998 (first entry)
 DE Human SRCR protein fragment.
 KM Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;
 KW nervous system; medullo-blastoma; glioma; breast; detection;
 KW autoantibody.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 3
 FT label= unknown
 PN W09830687-A2.
 PD 16-JUL-1998.
 PF 09-JAN-1998; D00096.
 PR 18-JUL-1997; DE-030997.
 PR 09-JAN-1997; DE-000519.
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 PI Mollenhauer J, Poustka A;
 DR WPI: 98-399136/34.
 DR N-PSDB; V49651.
 PT Proteins containing scavenger receptor, cysteine rich domain -
 PT useful for diagnosis and treatment of tumours
 PS Claim 1; Fig 1; 54pp; German.
 CC This sequence represents a fragment of a human protein which contains a
 CC SRCR (scavenger receptor, cysteine-rich) domain. The gene and encoded
 CC protein can be used to diagnose or treat tumours, particularly of the
 CC nervous system (medullo-blastoma or glioma) or breast. The DNA sequence
 CC and protein derived from it, are used to identify genes that express
 CC SRCR-domain containing proteins, to determine the form in which these
 CC proteins exist and to assess the significance of individual forms on
 CC cellular properties. The protein can be used to detect the presence of
 CC autoantibodies, and Ab which regulate its expression.
 SQ Sequence 666 AA;

Query Match 30.0%; Score 229; DB 1; Length 666;
 Best Local Similarity 38.2%; Pred. No. 2,22e-14;
 Matches 39; Conservative 16; Mismatches 45; Indels 2; Gaps 2;

DB 61 LRLVGGDRCRVRYLYRGSGVTCDDSDTNDANVYCRQIGCGWASAPGNARF-GQG 119
 117 VRLRGKNEFEIGVEYVAGVWGTVCSSHMDDSDASVICHOLQLG-KGIKOTPFSGLG 175

CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to growth
CC or survival of nerve cells including Parkinson's disease, Alzheimer's
CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
SQ Sequence 347 AA;

Query Match 29.5% Score 225; DB 1; Length 347;
Best Local Similarity 35.6% Pred. NO. 6.04e-14;
Matches 37; Conservative 20; Mismatches 43; Indels 4; Gaps 4;
DB 244 LRLVGGDNLCGRLEVLHKGWGSVCDDNMGEKEDQVCKQLGCGKSLSPFRDKCYGP 303
QY 117 VRLRGKNEFEGRVEYVASGVWGTVCSSHWDDSDASVICHQLQLG-GKGLA-KQTFPSGL 174
DB 304 GVGRI-WLDNVRCSGEGQSLDQCQHRFWGFDCTHOEDNAVICS 346
QY 175 GLPIYW-SNVRGRDDENTILCEKDIWGGVCPKMAAAVTCSS 217

Search completed: Mon Mar 13 10:25:46 2000
Job time : 11 secs.

THIS PAGE BLANK (OPTION)

334-433/SCRC3
Use last No. 5

WIDEVIEW (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h-pep protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 13 10:31:53 2000; MasPar time 6.67 Seconds
Tabular output not generated. 355.116 Million cell updates/sec

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.pep (5 of 6)
Perfect Score: 787
Sequence: 1 IRLAGGSGHSEGRLEVVYRG.....RROWGRHDCSHREDVSIACY 100

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 28.512; Variance 102.112; scale 0.279

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description	Pred. No.
787	100.0	822	1 W9087	Human serine protease	3.08e-78
787	100.0	875	1 W8361	Human neutrolypsin.	5.48e-67
691	87.8	761	1 W8362	Mouse neutrolypsin.	5.48e-67
691	87.8	761	1 W9908	Mouse serine protease	5.48e-67
412	52.4	451	1 W19708	Macrophage scavenger r	1.05e-34
407	51.7	451	1 R27035	Human sol. scavenger r	3.88e-34
407	51.7	451	1 R40802	Human scavenger recept	3.88e-34
407	51.7	451	1 W08077	Type I macrophage scav	3.88e-34
406	51.6	451	1 R2512	Human scavenger recept	5.05e-34
404	51.3	453	1 R05509	Scavenger receptor pro	8.53e-34
11	404	51.3	1 R27036	Bovine sol. scavenger	8.53e-34
12	363	46.1	1 R40172	Sequence of an immunor	3.82e-29
13	363	46.1	1 R40213	Sequence of a 90k tumo	3.82e-29
14	363	46.1	1 R36533	Gp85-97 clone 18 prod.	3.82e-29
15	363	46.1	1 W81363	Human cytokine regulat	3.82e-29
16	363	46.1	1 R41359	Tumour associated 90k	3.82e-29
17	362	46.0	1 W64590	Human SRCR protein fra	4.96e-29
18	362	46.0	1 W64591	Human SRCR protein.	4.96e-29
19	352	44.7	1 W39749	Mouse macroSR protein.	6.67e-28
20	352	44.7	1 W03561	Macrophage receptor wi	6.67e-28
21	351	44.6	1 W07609	Rat von Ebner's gland	8.65e-28
22	347	44.1	1 W39747	Human macroSR protein	2.44e-27
23	347	44.1	1 Y04374	Human macrophage recep	2.44e-27

24	347	44.1	520	1 W39748	Human macroSR protein	2.44e-27
25	343	43.6	347	1 W64537	Human liver cell clone	6.89e-27
26	343	43.6	347	1 W68200	Human scavenger recept	6.89e-27
27	343	43.6	347	1 Y13369	Amino acid sequence of	6.89e-27
28	339	43.1	574	1 R40784	Sequence of cyclophilin	1.94e-26
29	339	43.1	574	1 W81362	Mouse cytokine regulat	1.94e-26
30	105	13.3	349	1 W35849	Human CD5 for use in T	1.59e-01
31	96	12.2	356	1 W46917	Amino acid sequence of	1.08e+00
32	96	12.2	798	1 R57283	Bovine enterokinase.	1.08e+00
33	84	10.7	2509	1 W32881	Protein (OA-519) cross	1.27e+01
34	81	10.3	135	1 R28982	Thyroid N-acetyl-glucos	2.31e+01
35	78	9.9	337	1 W06579	Lipo-oligosaccharide g	4.16e+01
36	78	9.9	337	1 R91314	N. gonorrhoeae glycosyl	4.16e+01
37	78	9.9	872	1 W75912	Helicobacter valyl-L-TRN	4.16e+01
38	77	9.8	172	1 W72084	HSV-2 strain SB5 Conti	5.06e+01
39	77	9.8	172	1 W72210	HSV-2 strain SB5 Conti	5.06e+01
40	77	9.8	193	1 W72157	HSV-2 strain SB5 Conti	5.06e+01
41	77	9.8	335	1 R25063	Soluble human IL-5 rec	5.06e+01
42	77	9.8	420	1 R22215	Sequence of human inte	5.06e+01
43	77	9.8	420	1 R22219	Sequence of secretory	5.06e+01
44	77	9.8	874	1 W8698	H. pylori GHPD 686 pro	5.06e+01
45	77	9.8	1250	1 W59361	Drosophila melanogaste	5.06e+01

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	W9087	standard; Protein; 822 AA.
AC	W9087;	
DT	13-MAR-1999	(first entry)
DE	Human serine protease BSSP-3.	
KW	Serine protease; BSSP-3; brain tissue.	
OS	Homo sapiens.	
PN	W0905290-A1.	
PD	04-FEB-1999.	
PF	24-JUL-1998; J03324.	
PR	24-JUL-1997; JP-213699.	
PA	(SUNR) SUNTOXY LTD.	
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;	
DR	WPI: 99-142942/12.	
DR	N-PSDB; X19024.	
PT	New serine protease expressed in brain tissue - used in screening	
PT	for potential serine protease inhibitors for drug use	
PS	Claim 1; Page 61-65; 69pp; Japanese.	
CC	The present sequence is a serine protease designated BSSP-3, which	
CC	is isolated from human brain tissue. Transformants may be used to	
CC	produce the enzyme or its partial sequences. Products from the present	
CC	invention are used for screening for potential peptide or non-peptide	
CC	serine protease inhibitors or expression regulators for use as drugs.	
SO	Sequence 822 AA;	
Query Match	100.0%; Score 787; DB 1; Length 822;	
Best Local Similarity	100.0%; Pred. No. 3.08e-78;	
Matches	100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	334 IRLAGGSGHSEGRLEVVYRGWGTVCDDGWTLENTYVVCROLGPKYKXASAHFEESTG 393	
OY	334 IRLAGGSGHSEGRLEVVYRGWGTVCDDGWTLENTYVVCROLGPKYKXASAHFEESTG 393	
Db	394 PIMLDVSCSGKETRFLOCSRRQGRHDCSHREDVSIACY 433	
OY	394 PIMLDVSCSGKETRFLOCSRRQGRHDCSHREDVSIACY 433	
RESULT	2	
ID	W83361	standard; Protein; 875 AA.
AC	W83361;	
DT	17-FEB-1999	(first entry)
DE	Human neutrolypsin.	
KW	Human; neutrolypsin; tumour inhibition; neurological disease;	
KW	lung disease; gene therapy; drug development; stroke; brain injury;	
KW	neurodegeneration; neuroinflammatory disease; multiple sclerosis;	
KW	epilepsy; hypoxia; ischaemia; nerve transection; neovascularogenesis;	

KW Macrophage scavenger receptor protein; MSRP; human; lipoteichoic acid;
KW LTA; Gram-positive bacteria; cell wall; atherosclerosis; host defence;
KW septicæmia; complement activation; cytokine release; therapy;
KW nitric oxide production; bacterial infection; coptic check

QY 392 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 8

ID W08077: standard; Protein: 451 AA.

AC W08077:

DT 25-FEB-1997 (first entry)

DE Type I macrophage scavenger receptor.

KW Type I macrophage scavenger receptor; SRG; kidney; 293 cell;

KW cell substrate; cell attachment; atherosclerosis.

OS Homo sapiens.

PN W09638725-A1.

PD 05-DEC-1996.

PT 30-MAY-1995; US-453117.

PR (SMK) SMITHKLINE BEECHAM CORP.

PI Brawner ME, Elshourbagy MAE, Lysko PG;

WIPI: 97-034505/03.

DR N-PSDB: T43259.

AB New modified human embryonic kidney 293 cells - transfected with a human scavenger receptor gene to enhance ability to attach to a solid support.

PS Disclosure: Page 10-13; 32pp; English.

CC Human type I (W08077) and type II (W08078) macrophage scavenger receptors can be expressed in human embryonic kidney 293 cells following transfection of the cells with vectors carrying the respective scavenger receptor genes (T43259 and T43260). The transfected cells show an enhanced ability to attach to a solid support and are useful for screening cpds. for biological activity, or for identifying antagonists of the scavenger receptor gene, e.g. to develop agents for treatment of atherosclerosis.

CC Sequence 451 AA:

Query Match 51.7%; Score 407; DB 1; Length 451;

Best Local Similarity 51.5%; Pred. No. 3.88e-34;

Matches 52; Conservative 18; Mismatches 28; Indels 3; Gaps 2;

DB 350 VRLVGGSGPHEGRVETLHSGOMGTICDDRREVRVGVGCRSLGYP-GVQAVHKAHFGOG 408

QY 334 IRLAGKGSHEGRLEYYRGOMGTVCDDGTELTNTYVVCROLGFKYKQA--SANHEERS 391

DB 409 TGPIMLNEVFCFRESSIEECKRIRMGTRACSHSDAGVTC 449

QY 392 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 9

ID R21512: standard; Protein: 451 AA.

AC R21512:

DT 19-MAY-1992 (first entry)

DE Human scavenger receptor type I.

OS Homo sapiens.

PN J03290184-A.

PD 19-DEC-1991.

PT 06-APR-1990; 090274.

PR 06-APR-1990; JP-090274.

PA (CHUS) CHUGAI PHARMACEUTICAL KK.

WIPI: 92-051436/07.

DR N-PSDB: 021549.

PT New scavenger receptor-producing animal cells - which have been transformed with vector contg. gene for coding human scavenger receptor I or II type under control of promoter, etc.

PS Disclosure: Fig 1; 9pp; Japanese.

CC The gene encoding the receptor can be expressed from a cytomegalovirus promoter in a host cell e.g. CHO cell. The resulting recombinant scavenger receptor can be used to detect modified lipoproteins or modified substances in the blood.

CC See also R21513.

CC Sequence 451 AA:

Query Match 51.6%; Score 406; DB 1; Length 451;

Best Local Similarity 50.5%; Pred. No. 5.05e-34;

Matches 51; Conservative 19; Mismatches 28; Indels 3; Gaps 2;

DB 350 VRLVGGSGPHEGRVETLHSGOMGTICDDRREVRVGVGCRSLGYP-GVQAVHKAHFGOG 408

QY 334 IRLAGKGSHEGRLEYYRGOMGTVCDDGTELTNTYVVCROLGFKYKQA--SANHEERS 391

DB 409 TGPIMLNEVFCFRESSIEECKRIRMGTRACSHSDAGVTC 449

QY 392 TGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 10

ID R05509: standard; protein: 453 AA.

AC R05509;

DT 23-OCT-1990 (first entry)

DE Scavenger receptor protein with affinity for acylated low density

DE lipoprotein (LDL).

KW Acylated low density lipoprotein; LDL; atherosclerotic plaque;

OS ds

PN W09005748-A.

PD 31-MAY-1990.

PT 14-NOV-1989; 005116.

PR 15-NOV-1988; US-272002.

PA (MAST) Massachusetts Inst Tech.

PI Engellennet TT.

DR WPI: 90-193408/25.

DR N-PSDB: 004926.

PT New receptor protein - has affinity for acetylated low density lipoprotein and corresponding antibodies and DNA sequences.

PS Disclosure: 79pp; English.

CC Receptor protein, and fragments thereof may be immobilised on a support and used in assay and purification of the CC LDL target. Labelled Abs, raised to the protein may be injected into the vascular system to detect the presence of atherosclerotic CC plaques.

CC Sequence 453 AA:

Query Match 51.3%; Score 404; DB 1; Length 453;

Best Local Similarity 52.0%; Pred. No. 8.53e-34;

Matches 53; Conservative 19; Mismatches 25; Indels 5; Gaps 4;

DB 352 VRLVGGSGPHEGRVETLHSGOMGTVCDDRW-ELRGVLVCRSLGK-GVQSHKAAHYGK 409

QY 334 IRLAGKGSHEGRLEYYRGOMGTVCDDGTELTNTYVVCROLGFKYKQASANH--FEE 390

DB 410 TGPIMLNEVFCFRESSIEECKRIRMGTRACSHSDAGVTC 451

QY 391 STGPIMLDVSCSGKETRFLQCSRROMGRHDCSHREDVSIAC 432

RESULT 11

ID R27036: standard; Protein: 453 AA.

AC R27036;

DT 16-FEB-1993 (first entry)

DE Bovine sol. scavenger receptor.

KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.

OS Bos taurus.

FH Key

FT domain Location/Qualifiers

FT domain 1..50

FT domain /note="cytoplasmic"

FT domain 51..76

FT domain /note="transmembrane"

FT domain 77..109

FT domain /note="spacer"

FT domain 110..271

FT domain /note="alpha helical coiled coil"

FT domain 272..343

FT domain /note="collagen binding"

FT domain 341..451

FT domain /note="Cys-rich"

PN W09214482-A.

PD 03-SEP-1992.


```

CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 585 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 585 AA; 65330 MW; 1874129 CN;
SQ
Query Match          46.1%; Score 363; DB 1; Length 585;
Best Local Similarity 43.0%; Pred.No. 2.08e-27;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1.

Db      24 MRLDGGATNGRGVEIFRGQGWGVCDNLMDLTASVYCRALGPEMNTOLGRAFGQGS 83
QY      334 IRLGGKSHSGHRELVIRGVGVCDGDDGTELTITVYCRDLGFKYGRQA-SANHFEST 392
Db      84 GPIMLDEVQCTGTETASLADCKSLGMLSNCRHERDAGVVC 123
QY      393 GPIMLDVSCTSGKTRFLQCSRGRMGRRDCHRDVSLAC 432

RESULT      8
ID          US-08-316-714-10      STANDARD;      PRT;      585 AA.
AC          xxxxxx
DT
DX
DE
XX          Sequence 10, Application US/08316714
CC
CC          Sequence 10, Application US/08316714
CC          Patent No. 5965382
CC          GENERAL INFORMATION:
CC          APPLICANT: Kochs, Kirston E.
CC          APPLICANT: Halenbeck, Robert F.
CC          APPLICANT: Taylor, Eric W.
CC          APPLICANT: Wang, Alice M.
CC          APPLICANT: Casipit, Clayton L.
CC          TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC          NUMBER OF SEQUENCES: 4
CC          CORRESPONDENCE ADDRESSES:
CC          ADDRESSEE: Cetus Oncology Corporation
CC          STREET: 1400 Fifty-Third Street
CC          CITY: Emeryville
CC          STATE: CA
CC          COUNTRY: USA
CC          ZIP: 94608
CC          COMPUTER READABLE FORM:
CC          MEDIUM TYPE: Floppy disk
CC          COMPUTER: IBM PC compatible
CC          OPERATING SYSTEM: PC-DOS/MS-DOS
CC          SOFTWARE: PatentIn Release #1.0, Version #1.25
CC          CURRENT APPLICATION DATA:
CC          APPLICATION NUMBER: US/08/316,714
CC          FILING DATE:
CC          CLASSIFICATION: 435
CC          PRIOR APPLICATION DATA:
CC          APPLICATION NUMBER: US/07/961,404
CC          FILING DATE: 15-OCT-1992
CC          ATTORNEY/AGENT INFORMATION:
CC          NAME: Goldman, Kenneth M.
CC          REGISTRATION NUMBER: 34,174
CC          REFERENCE/DOCKET NUMBER: 2595.1
CC          TELECOMMUNICATION INFORMATION:
CC          TELEPHONE: (510) 420-3152
CC          TELEFAX: (510) 658-5470
CC          TELEX: N/A
CC          INFORMATION FOR SEQ ID NO: 10:
CC          SEQUENCE CHARACTERISTICS:
CC          LENGTH: 585 amino acids
CC          TYPE: amino acid
CC          STRANDEDNESS: single
CC          TOPOLOGY: linear

```


CC MOLECULE TYPE: protein
SQ SEQUENCE 585 AA; 65330 MW; 1874129 CN;
Query Match 46.1%; Score 363; DB 2; Length 585;
Best Local Similarity 43.0%; Pred. No. 2,08e-27;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;
DB 24 MRLADGATNGRVEIFRYGQWGVCDNLMDLTDASVVCALGFENATQALGRAAFQGS 83
Y 334 IRLAGKSGSHGRLVYRGQWGVCDGWTGLNTYVVCRLQGFYKQKA-SANHFEEST 392
DB 84 GPIMLDEVQCTGTASLADCKSLGMLKSNCRHERDAGVC 123
Y 393 GPIWLDVSCSGKETRFLQCSRRQWRHDCSHREDVSIAC 432
T 9
US-08-477-674-10 STANDARD: PRT; 585 AA.
AC xxxxxx
DT
XX
XX
DE Sequence 10, Application US/08477674
CC
CC Sequence 10, Application US/08477674
CC Patent No. 5644035
CC GENERAL INFORMATION:
CC APPLICANT: Kochs, Kirston E.
CC APPLICANT: Halenbeck, Robert F.
CC APPLICANT: Taylor, Eric W.
CC APPLICANT: Wang, Alice M.
CC APPLICANT: Casipit, Clayton L.
CC TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cetus Oncology Corporation
CC STREET: 1400 Fifty-third street
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/477,674
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/316,714
CC FILING DATE:
CC APPLICATION NUMBER: US/07/961,404
CC FILING DATE: 15-OCT-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Goldman, Kenneth M.
CC REGISTRATION NUMBER: 34,174
CC REFERENCE/DOCKET NUMBER: 2595.1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (510) 420-3152
CC TELEFAX: (510) 658-5470
CC TELEX: N/A
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA; 65330 MW; 1874129 CN;

Query Match 46.1%; Score 363; DB 1; Length 585;
Best Local Similarity 43.0%; Pred. No. 2,08e-27;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;
DB 24 MRLADGATNGRVEIFRYGQWGVCDNLMDLTDASVVCALGFENATQALGRAAFQGS 83
Y 334 IRLAGKSGSHGRLVYRGQWGVCDGWTGLNTYVVCRLQGFYKQKA-SANHFEEST 392
DB 84 GPIMLDEVQCTGTASLADCKSLGMLKSNCRHERDAGVC 123
Y 393 GPIWLDVSCSGKETRFLQCSRRQWRHDCSHREDVSIAC 432
RESULT 10
ID US-08-794-795-7 STANDARD: PRT; 489 AA.
AC xxxxxx
DT
XX
XX
DE Sequence 7, Application US/08794795
CC
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T.
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 489 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 489 AA; 49441 MW; 1141296 CN;
Query Match 44.7%; Score 352; DB 2; Length 489;
Best Local Similarity 42.4%; Pred. No. 2,99e-26;
Matches 42; Conservative 23; Mismatches 30; Indels 4; Gaps 3;
DB 394 VRIMGFT-NR-GRAEYVYNNEMGTICDDMDNNDATVFCMLGLSGNRALSS--YGGGSG 449

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA; 49764 MW; 1218428 CN;
Query Match 44.1%; Score 347; DB 2; Length 495;
Best Local Similarity 44.4%; Pred. No. 1.00e-25;
Matches 44; Conservative 21; Mismatches 30; Indels 4; Gaps 3;
DB 399 VRIVGS-SNR-GRAEYVYSTGTICDDWQNSDAIVFCRMIG--YSKGRALYKVGAGTG 454
QY 334 IRLAGGKSGHEGRLEYYRQWGTVCDDGWTEINTYVVCRLGFKYKQASANHFEESTG 393
DB 455 QIWLNVQCRGTESTLMSCTKNSMGHHDCSHEDAGVEC 493
QY 394 PIVLDVSCSGKRETRFLQCSRQWRGHRHDCSHREDVSIAC 432
RESULT 14
ID US-08-794-795-6 STANDARD; PRT: 520 AA.
XX
AC xxxxxx
XX
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:

CC APPLICANT: Elshourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lysko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Smithkline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA; 52638 MW; 1340662 CN;
Query Match 44.1%; Score 347; DB 2; Length 520;
Best Local Similarity 44.4%; Pred. No. 1.00e-25;
Matches 44; Conservative 21; Mismatches 30; Indels 4; Gaps 3;
DB 424 VRIVGS-SNR-GRAEYVYSTGTICDDWQNSDAIVFCRMIG--YSKGRALYKVGAGTG 479
QY 334 IRLAGGKSGHEGRLEYYRQWGTVCDDGWTEINTYVVCRLGFKYKQASANHFEESTG 393
DB 480 QIWLNVQCRGTESTLMSCTKNSMGHHDCSHEDAGVEC 518
QY 394 PIVLDVSCSGKRETRFLQCSRQWRGHRHDCSHREDVSIAC 432
RESULT 15
ID US-08-681-151-1 STANDARD; PRT: 356 AA.
XX
AC xxxxxx
XX
DE Sequence 1, Application US/08681151
XX
CC Sequence 1, Application US/08681151
CC Patent No. 5869637
CC GENERAL INFORMATION:
CC APPLICANT: Au-Young, Janice
CC APPLICANT: Bandman, Olga
CC APPLICANT: Braxton, Scott Michael
CC APPLICANT: Goli, Surya
CC TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN

(天)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

    hpp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Mar 13 10:31:24 2000;      MasPar time 10.32 Seconds
Tabular output not generated.              457.141 Million cell updates/sec

```

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.ppt (5 of 6)
Perfect Score: 987
Sequence: 1 IRLGGGSGSHGRLVYRQ.....RQQRHDCSHREDVSTACY 100

Scoring table: PAM 150

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62

Statistics: Mean 37.820; Variance 61.366; scale 0.616

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	691	87.8	761	2	JC5759		brain-specific serine	2.55e-14
2	419	53.2	918	2	JC4361		scavenger receptor Cy	2.42e-75
3	413	52.5	453	2	S08276		LDL receptor I, macro	7.09e-74
4	412	52.4	1116	2	S36077		M130 antigen - human	1.24e-73
5	412	52.4	1149	2	I38006		M130 antigen (extract	1.24e-73
6	412	52.4	1151	2	I38004		M130 antigen (cytosol	1.24e-73
7	412	52.4	1156	2	I38005		M130 antigen (cytosol	1.24e-73
8	412	52.4	2155	2	I14895		scavenger receptor cy	1.24e-73
9	407	51.7	451	2	A38415		macrophage scavenger	2.07e-72
10	405	51.5	454	2	I46862		macrophage scavenger	6.36e-72
11	397	50.4	454	2	A4407		macrophage scavenger	5.65e-70
12	393	49.9	458	2	B44407		macrophage scavenger	5.31e-69
13	391	49.7	1435	2	A46496		antigen M61.1 precurs	1.63e-68
14	374	47.5	600	2	S55744		mucln (clone pGM-1)	2.15e-64
15	363	46.1	585	2	A4161		Mac-2-binding glycop	9.75e-62
16	352	44.7	518	2	A55840		macrophage bacteria-b	4.33e-59
17	351	44.6	1290	2	A57190		ebherin precursor - r	7.53e-59
18	350	44.5	577	2	A53202		Cytophilin C-associa	1.31e-58
19	346	44.0	504	2	S26745		mucln (clone pGM1-1)	1.19e-57
20	293	37.2	468	2	S26741		T-cell glycoprotein C	4.58e-45
21	281	35.7	532	2	A33751		spectac receptor prec	2.95e-42
22	281	35.7	626	2	I45100		mscd6 precursor - mu	2.95e-42
23	114	14.5	495	1	S11270		T-cell surface glyco	3.35e-06

45	106	13.5	1034	A53663	entereptelidase (EC 3	9.57e-05
24	105	13.3	495	A26396	T-cell surface glycop	1.44e-04
25	106	13.5	1034	A53663	entereptelidase (EC 3	1.61e-03
26	99	12.6	442	A47074	gene CD5 protein sh	1.61e-03
27	99	12.6	1019	A56318	entereptelidase (EC 3	5.25e-03
28	96	12.2	1035	A43030	entereptelidase (EC 3	1.67e-02
29	93	11.8	558	T15448	hypothetical protein	2.45e-02
30	92	11.7	202	T01605	blue copper-binding p	3.58e-02
31	91	11.6	613	S15468	complement C3b/C4b in	3.58e-02
32	91	11.6	2616	A57096	nudel protein precurs	5.21e-02
33	90	11.4	501	S43536	T-cell surface glycop	7.57e-02
34	89	11.3	2206	U25280	voltage-dependent cal	1.10e-01
35	88	11.2	333	S15257	hypothetical protein	1.10e-01
36	88	11.2	494	A29079	lymphocyte surface gl	1.10e-01
37	88	11.2	1113	JEO313	low-density lipoprote	2.29e-01
38	86	10.9	583	A29154	complement factor I (4.72e-01
39	84	10.7	2504	A57768	fatty-acid synthase (6.75e-01
40	84	10.7	2509	G01880	probable acyl-CoA deh	6.75e-01
41	83	10.5	411	C69534	probable DNA helicase	9.61e-01
42	83	10.5	670	G12121	mannose-specific lect	9.61e-01
43	82	10.4	149	A58801	hypothetical protein	9.61e-01
44	82	10.4	306	S50689	conserved hypotetica	9.61e-01
45	82	10.4	340	C71266		

ALIGNMENTS

[illegible]


```
921-1024          #domain scavenger receptor cysteine-rich domain homology
                    #label SRC9
SUMMARY           #length 1116 #molecular-weight 120979 #checksum 5687

Query Match      52.4% Score 412; DB 2; Length 1116;
Best Local Similarity 48.5%; Pred. No. 1,24e-73;
Matches 49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;

Db    714 LRLVNGGRCACAGVELIYHESGWGITCDDSW-DLSDAHVCROLGGCAINATGSANHGEG 772
       ||| | | | | | | | | | | | | | : : : | | | | : : : | | |
Ox   334 IRLAGKSGSHEGRLEYYRQMGCTVDCDGTTEL-NRTWVCROLGFKYKGASAN-HFEES 391
       :|| | | | | | | | | | | | | | : : | | | | | : : | | |

Db    773 TGPIMLDEMKCNKESRIMOCHSHGGOONCRKHEDAGVIC 813
       |||||||::|:|::|::|:|:|:|::|:|:~::~|:|:~::~|
Ox   392 TGPIMLDIVSCSGKETRFLOCSRROWRHCDPSHREDYSIAC 432
       |||||||::|:|::|::|:|:|:|::|:|:~::~|:|:~::~|

#status     #type complete
#accession  138006
#authors     Lewis, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
              Willis, A.C.; Mason, D.Y.
#journal     Eur. J. Immunol. (1993) 23:2320-2325
              A new macrophage differentiation antigen which is a member of
              the scavenger receptor superfamily.
#cross-references MWID:93380506
#accession  138005 preliminary; translated from GB/EMBL/DDBJ
#status     ##molecule_type mRNA
              ##residues 1-1149 #label RES
              ##cross-references EMBL:Z22971; NDB:g312147; PID:g312148
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE        43-147
               151-254 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC1\
               258-361 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC2\
               365-468 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC3\
               ~0-573 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC4\
               ~8-711 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC5\
               744-847 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC6\
               849-953 #domain scavenger receptor cysteine-rich domain homology
                   #label SRC7\
               954-1057 #domain scavenger receptor cysteine-rich domain homology
                       #label SRC8\
                       #domain scavenger receptor cysteine-rich domain homology
                           #label SRC9
SUMMARY         #length 1149 #molecular-weight 124328 #checksum 487

Query Match      52.4% Score 412; DB 2; Length 1149;
Best Local Similarity 48.5%; Pred. No. 1,24e-73;
Matches 49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;
```

TITLE	M130 antigen (cytosolic variant 1) - human
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997
ACCESSIONS	I38004; S36078
REFERENCE	I38003
#authors	Lau, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
#journal	Eur. J. Immunol. (1993) 23:2320-2325
#title	A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#cross-references MUID:	93380506
#accession	I38004
#status	preliminary: translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-1151 #label RES
#cross-references EMBL:	Z22970; NID:g312143; PID:g312144
CLASSIFICATION	#superfamily scavenger receptor cysteine-rich domain homology cytosol
FEATURE	
43-147	#domain scavenger receptor cysteine-rich domain homology #label SRC1\
151-254	#domain scavenger receptor cysteine-rich domain homology #label SRC2\
258-361	#domain scavenger receptor cysteine-rich domain homology #label SRC3\
365-468	#domain scavenger receptor cysteine-rich domain homology #label SRC4\
470-573	#domain scavenger receptor cysteine-rich domain homology #label SRC5\
575-678	#domain scavenger receptor cysteine-rich domain homology #label SRC6\
711-814	#domain scavenger receptor cysteine-rich domain homology #label SRC7\
816-920	#domain scavenger receptor cysteine-rich domain homology #label SRC8\
921-1024	#domain scavenger receptor cysteine-rich domain homology #label SRC9
SUMMARY	#length 1151 #molecular_weight 124820 #checksum 1481
Query Match	52.4% Score 412; DB 2; Length 1151;
Best Local Similarity	48.5%; Pred. NO.1.24e-73;
Matches	49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;
Dd	714 LRLVGGRCAGREYIHESGWTICDDSD-DLSDDHYVCROLGCGEALNATGSAHFEGG 772 :: :: :: :: :: :: :: :: :: :: Oy 334 IRLAGGKSHHGRLRYVRGMQKVCDDEWTEL-NIVYVCROLDGRKYGRKASAN-HFEBS 391
Dd	773 TGPIMLDEMCKNGEKSRIRWOCHSGHGGQGNCRRHKEDAGVIC 813 :: :: :: :: :: :: :: :: Oy 392 TGPIMLDVSCSGKETRFLLQCSRROGMRDCHSHREDVSIAIC 432
RESULT	7
ENTRY	I38005 #type complete
TITLE	M130 antigen (cytosolic variant 2) - human
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997
ACCESSIONS	I38005; S36079
REFERENCE	I38003
#authors	Lau, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.
#journal	Eur. J. Immunol. (1993) 23:2320-2325
#title	A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.
#cross-references MUID:	93380506
#accession	I38005
#status	preliminary: translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-1156 #label RES
#cross-references EMBL:	Z22970; NID:g312145; PID:g312146
CLASSIFICATION	#superfamily scavenger receptor cysteine-rich domain homology

KEYWORDS	FEATURE	cytosol
43-147	#domain scavenger receptor cysteine-rich domain homology #label SRC1\	
151-254	#domain scavenger receptor cysteine-rich domain homology #label SRC2\	
258-361	#domain scavenger receptor cysteine-rich domain homology #label SRC3\	
365-468	#domain scavenger receptor cysteine-rich domain homology #label SRC4\	
470-573	#domain scavenger receptor cysteine-rich domain homology #label SRC5\	
575-678	#domain scavenger receptor cysteine-rich domain homology #label SRC6\	
711-814	#domain scavenger receptor cysteine-rich domain homology #label SRC7\	
816-920	#domain scavenger receptor cysteine-rich domain homology #label SRC8\	
921-1024	#domain scavenger receptor cysteine-rich domain homology #label SRC9	
Query Match	52.4% Score 412; DB 2; Length 1156;	
Best Local Similarity	48.5% Pred. NO. 1.24e-73;	
Matches	49; Conservative 25; Mismatches 24; Indels 3; Gaps 3;	
Db	714 IRLVGGRCAGKAVETLYHEGSMGTICDDSN-DISDAVVCROLGCGAINTGSAHFEGG 772	
Qy	334 IRLAGKSGHEGLLEVYRRQMGTVCDGDWTEL-NTYVVCROLGFKYKQKASAN-HFEES 391	
Db	773 TGPVIMDEKCKNGESRIMQCHSHGQNCQCKREDAVGLC 813	
Qy	392 TGPVIMLDVSCSGKETRFLOCSRRQGRHDCSHREDVSIAC 432	
RESULT	8	
ENTRY	T14893	#type complete
TITLE	scavenger receptor cysteine-rich protein precursor - sea urchin (Strongylocentrotus purpuratus)	
ORGANISM	#formal_name Strongylocentrotus purpuratus #common_name purple urchin	
DATE	20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999	
ACCESSIONS	T14893	
REFERENCE	Z18253	
#authors	Panzer, Z.; Rast, J.P.; Davidson, E.H.	
#journal	Immunogenetics (1999) 49:773-786	
#file	Origins of immunity: transcription factors and homologs of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes.	
cession	T14893	
#status	preliminary; translated from GB/EMBL/DBJ	
#molecule_type	mRNA	
#residues	1-2153 #label PAN	
#cross-references	EMBL:AF064259; NID:g4165052; PID:g4165053; PIDN:AD08654.1	
GENETICS		
#gene	SRCR12	
SUMMARY	#length 2153 #molecular-weight 226556 #checksum 3510	
Query Match	52.4% Score 412; DB 2; Length 2153;	
Best Local Similarity	48.0% Pred. No. 1.24e-73;	
Matches	48; Conservative 25; Mismatches 26; Indels 1; Gaps 1;	
Db	664 MRLVGGNSRQGRLEISINNQMGTVCDSDINDATVVCROLGSSAVSAFTSAHFQGS 743	
Qy	334 IRLAGKSGHEGLLEVYRRQMGTVCDGDWTELNTYVVCROLGFKYK-KASANHFEST 392	
Db	744 GTIWLDDVSCAGNENSLMDGCHGLGVHNAHMDADEV 783	
Qy	393 GTIWLDDVSCSGKETRFLOCSRRQGRHDCSHREDVSIAC 432	

```

RESULT      9                                     #type complete
ENTRY       A38415                               macrophage scavenger receptor splice form I - human
TITLE       #format_name Homo sapiens; #common_name man
ORGANISM    28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change
DATE        20-Sep-1999
ACCESSIONS  A38415; A44408
REFERENCE   Masumoto, A.; Naito, M.; Itakura, H.; Ikemoto, S.; Asaoka,
#authors    H.; Hayakawa, I.; Kanamori, H.; Aburatani, H.; Takahashi, F.;
            Suzuki, H.; Kobari, Y.; Miyai, T.; Takahashi, K.; Cohen,
            E.H.; Wydro, R.; Hausman, D.E.; Kodama, T.
            Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9133-9137
            Human Macrophage Scavenger Receptors: Primary Structure,
            Expression, and Localization in Atherosclerotic Lesions.
            PMID:91067661
#cross-references MIMD:91067661
#accesion     A38415
##status      Preliminary
##molecule_type mRNA
##residues    1-451 ##label MAT
#cross-references GB:D90187; NID:g219989; PIDN:BAAJ4208.1;
                PUD:d1014913; PID:g219990
REFERENCE     A44408
#authors      EmI, M.; Asaoka, H.; Matsumoto, A.; Itakura, H.; Kurihara,
              Y.; Made, T.; Kanamori, H.; Izaki, Y.; Takahashi, E.;
              Lepert, M.; Lalouel, J.; Kodama, T.; Mukai, T.
              J. Biol. Chem. (1993) 268:2120-2125
              Structure, Organization, and Chromosomal Mapping of the Human
              Macrophage Scavenger Receptor Gene.
#cross-references MIMD:93131971
#accesion     A44408
##status      preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues    155-272 ##label EX1
#note          sequence extracted from NCBI backbone (NCBIP:123189)
GENETICS
#gene         GDB:MSR1
##cross-references GDB:128046; OMTM:153622
CLASSIFICATION
#map_position 8p22-8q22
               #superfamily unassigned collagens; scavenger receptor
               cysteine-rich domain homology
KEYWORDS      alternative splicing; coiled coil; transmembrane protein
FEATURE
347-450       #domain scavenger receptor cysteine-rich domain homology
               #label SRC
SUMMARY       #length 451 #molecular_weight 49762 #checksum 6168
               #length 451; Score 407; DB 2; Length 451;
               Best Local Similarity 51.5%; Pred. No. 2.07e-72;
               Matches 52; Conservative 18; Mismatches 28; Indels 3; Gaps 2;
Db           350 VRLVGGSGPHGREGVEILHSCGWGITCDDRREVVAGGVYCRSLGYP-GVOAYVKRAAHFGOG 408
             :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Oy           334 IRLVGKGSHSGRGRELVYVRGOWGTCCDDGETLNTVTVYRDLGFKKYGQA--SANHFES 391
Db           409 TGPIWLNEVFEGRESSIECKTRQWCTKRSCHSEADGYTC 449
             ||||||::::|:::||::||::||::||::||::||::||::||::||::||::||::||:
Oy           392 TGPIWLDDVSCSGKETRFLOCSRQRMRHCSDREDYSTAAC 432

RESULT      10
ENTRY       I46862                               type complete
TITLE       macrophage scavenger receptor type I - rabbit
ORGANISM    #format_name Oryctolagus cuniculus; #common_name domestic
            rabbit
DATE        14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change
ACCESSIONS  I46862
REFERENCE   Bickel, P.E.; Freeman, M.W.
            J. Clin. Invest. (1992) 90:1450-1457
            Rabbit aortic smooth muscle cells express inducible
            macrophage scavenger receptors messenger RNA that is absent

```



```

TITLE      macrophage scavenger receptor, MSRI - mouse
ALTERNATE_NAMES
ORGANISM   macrophage scavenger receptor type I
DATE       #formal_name Mus musculus #common_name house mouse
          30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
          20-Sep-1999
ACCESSIONS BAA4407; A38260; I56334
REFERENCE  #authors
          #journal
          #title
#cross-references MUID:93131972
#accession    BAA407
##status     preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues   1-458 ##label DOI
##experimental_source macrophage-like cell line P38D
#note        sequence extracted from NCBI backbone (NCBIP:123207)
REFERENCE  #authors
          #journal
          #title
#cross-references MUID:91062370
#accession    A38260
##status     preliminary
##molecule_type mRNA
##residues   349-458 ##label FRE
#cross-references GB:M59445; GB:N36817; NID:g192737; PIDN:AAA37464.1;
          PID:g192738
REFERENCE  #authors
          #journal
          #title
I56334
Ashkenas, J.; Penman, M.; Vasile, E.; Acton, S.; Freeman,
M.W.; Krieger, M.
J. Lipid Res. (1993) 34:983-1000
Structures and high and low affinity ligand binding
properties of murine type I and type II macrophage
scavenger receptors.
#cross-references MUID:93359822
#accession    I56334
##status     preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues   5-458 ##label RES
#cross-references GB:L04574; NID:g293745; PIDN:AAA39747.1; PID:g293746
CLASSIFICATION #superfamily1 unassigned collagens; scavenger receptor
               cysteine-rich domain homology
FEATURE
354-457      #domain scavenger receptor cysteine-rich domain homology
SUMMARY      #length 458 #molecular_weight 50130 #checksum 1435
Query Match  49.9%; Score 393; DB 2; Length 458;
Best Local Similarity 47.0%; Pred. No. 5,31e-69;
Matches 477; Conservative 18; Mismatches 34; Indels 1; Gaps 1;
Db 357 VRLVGGSGAHEGRVEYFHQGWGTICDDRWDIRAGGVCRSLGYOEVLAVKRAHFGOGT 416
      :|||::|:||||::|:|||||::|:|||||::|:|||||::|:|||||::|:|||||::|:
Oy 334 IRLAGKGSHGHRLEVVYRGOMGTVCDDCWTELTNTFVVRQLGFKYGQA-SANFEEST 392
      |||||::|:|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Oy 393 GPIWLDVSCSGKETRFLOCSRRWRGHDCSHREDVSIAIC 432
RESULT      13
ENTRY
TITLE      A46496 #type complete
ORGANISM   antigen WCL1 precursor - bovine
DATE       #formal_name Bos primigenius taurus #common_name cattle
          18-Jun-1993 #sequence_revision 19-May-1994 #text_change

```



```

##cross-references GB:X79089; NID:9403473; PIDN:CAAS5699.1; PID:9483474
#32384
REFERENCE
#authors Iacobelli, S.; Buccell, I.; d'Egidio, M.; Giuliani, C.; Natoli, C.; Tinari, N.; Rubinstein, M.; Schlesinger, J.
#journal FEBS Lett. (1993) 319:59-65
#title Purification and characterization of a 90 kDa protein released from human tumors and tumor cell lines.
#cross-references MUID:93202277
#accession S32384
##molecule_type protein
#residues 19-24, 'C', 26-40 ##label IAC
REFERENCE
#authors Inohara, H.; Raz, A.
#journal Biochem. Biophys. Res. Commun. (1994) 201:1366-1375
#title Identification of human melanoma cellular and secreted ligands for galectin-3.
#cross-references MUID:94296411
#accession PC2211
##molecule_type protein
#residues 19-38 ##label INO
#note the reported molecular weight of this protein was 98K
#accession PC2212
##molecule_type protein
#residues 19-26 ##label IN2
#note the reported molecular weight of this protein was 70K
REFERENCE
#authors A41005
#journal Rosenberry, I.; Cherayil, B.J.; Isselbacher, K.J.; Pillai, S.
#title J. Biol. Chem. (1991) 266:18731-18736
#title Mac-2-binding glycoproteins. Putative ligands for a cytosolic beta-galactoside lectin.
#cross-references MUID:92011634
#accession A41005
##molecule_type protein
#residues 19-22, 'G', 24-27, 'XX', 30 ##label ROS
#note the reported molecular weight of this protein, designated Mac-2-binding glycoprotein 2 (M2BP-2), was 70K; the amino-terminal sequence of a related form of 98K, designated M2BP-1, could not be determined
#suprafamily scavenger receptor cysteine-rich domain homology extracellular scavenger receptor glycoprotein
CLASSIFICATION
KEYWORDS #domain signal sequence #struts predicted #label SIG
FEATURE #domain scavenger receptor cysteine-rich domain homology
1-18 #label SRC
21-24 #molecular-weight 65330 #checksum 2263
SUMMARY
ry Match 46.1%; Score 363; DB 2; Length 585;
t Local Similarity 43.0%; Pred. No. 9,75e-62;
Ches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;
Db 24 MRLADGATNOGRVEIYRGMGTVCNDLMDLTDSAVCRALGFENATDLAGRAFGGS 83
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |
QY 334 IRLAGGSGSHGRLEIYRGMGTVCDDGTETLNTIYVRCOLGFGYGRQA-SANHFEST 392
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |
QY 393 GPIWLDVSCGKTRFLQCSRRMGWRHDCSHEDVSIAC 432
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |

```

Search completed: Mon Mar 13 10:31:36 2000
Job time : 12 secs.

THIS PAGE BLANK (USPTO)

 WISE (TM)

Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:30:22 2000; Maspar time 6.44 Seconds
 Tabular output not generated. 463,783 Million cell updates/sec

Title: >US-09-147-947-6
 Description: A (334-433) from US09147947A.pep (5 of 6)
 Perfect Score: 787
 Sequence: 1 IRLAGKSGHEGRLEVVYRG.....RQGRHDCSHREDVSIACY 100

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 38.688; Variance 57.261; scale 0.676

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length DB	ID	Description	Pred. No.
787	100.0	875	1	NETR_HUMAN NEUTROTYPIN PRECURSOR	1.39e-183
691	87.8	761	1	NETR_MOUSE NEUTROTYPIN PRECURSOR	7.85e-157
413	52.5	433	1	MSRE_BOVIN MACROPHAGE SCAVENGER R	7.34e-81
407	51.7	451	1	MSRE_HUMAN MACROPHAGE SCAVENGER R	2.93e-79
405	51.5	454	1	MSRE_RABIT MACROPHAGE SCAVENGER R	1.00e-78
393	49.9	458	1	MSRE_MOUSE MACROPHAGE SCAVENGER R	1.56e-75
391	49.7	1436	1	MC1L_BOVIN ANTIGEN MC1.1	5.32e-75
293	37.2	468	1	CD6_HUMAN T-CELL DIFFERENTIATION	2.22e-49
281	35.7	532	1	SPER_STRPU EGG PEPTIDE SPERACTIN	2.60e-46
114	14.5	495	1	CD5_BOVIN T-CELL SURFACE GLYCOP	4.72e-07
106	13.5	1034	1	ENTR_PIG ENTEROPEPTIDASE PRECUR	1.75e-05
105	13.3	495	1	CD5_HUMAN T-CELL SURFACE GLYCOP	2.72e-05
100	12.7	492	1	TMS2_HUMAN TRANSMEMBRANE PROTEASE	2.40e-04
100	12.7	1069	1	ENTR_MOUSE ENTEROPEPTIDASE (EC 3	2.40e-04
99	12.6	1019	1	ENTR_HUMAN ENTEROPEPTIDASE PRECUR	3.68e-04
96	12.2	1035	1	ENTR_BOVIN ENTEROPEPTIDASE PRECUR	1.31e-03
91	11.6	2616	1	NDL_DROME SERINE PROTEASE NDUCL	1.04e-02
90	11.4	491	1	CD5_RAT T-CELL SURFACE GLYCOP	1.55e-02
88	11.2	494	1	CD5_MOUSE T-CELL SURFACE GLYCOP	3.46e-02
86	10.9	583	1	CEAL_HUMAN COMPLEMENT FACTOR I PR	7.62e-02
84	10.7	2504	1	FAS_HUMAN FATTY ACID SYNTHASE (E	1.66e-01
82	10.4	306	1	YE16_YEAST HYPOHETICAL 35.9 KD P	3.55e-01
80	10.2	756	1	RIRI_HABIN RIBONUCLEOSIDE-DIPHOS	7.51e-01

RESULT	1	STANDARD	PRT	875 AA	ALIGNMENTS
ID	NETR_HUMAN				
AC	P56730;				
DT	15-DEC-1999 (Rel. 39, Created)				
DT	15-DEC-1999 (Rel. 39, Last sequence update)				
DT	15-DEC-1999 (Rel. 39, Last annotation update)				
DE	NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN).				
GN	PRSS12.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
NC	Eutheria; Primates; Catarrhini; Homidae; Homo.				
RP	SEQUENCE FROM N.A.				
RX	TISSUE-BRAIN:				
RA	MEDLINE: 96201705.				
RT	PROBA K., GSCHWEND T.P., SONDERGGER P.;				
RL	"Cloning and sequencing of the cDNA encoding human neutrotypin.";				
CC	Biochim. Biophys. Acta 1396:143-147 (1998).				
CC	- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.				
CC	- SIMILARITY: CONTAINS 4 SRCR DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL: AJ001531; CA04816.1; -				
DR	PROSITE: PS00134; TRYPSIN_HIS. 1.				
DR	PROSITE: PS00135; TRYPSIN_SER. 1.				
DR	PROSITE: PS00420; SPERACTIN_RECEPTOR. 3.				
KW	Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.				
FT	SIGNAL	1	20	POTENTIAL.	
FT	CHAIN	21	875	NEUTROTYPIN.	
FT	DOMAIN	23	92	PROLINE-RICH.	
FT	DOMAIN	93	165	KRINGLE.	
FT	DOMAIN	170	271	SRCR 1.	
FT	DOMAIN	280	381	SRCR 2.	
FT	DOMAIN	387	487	SRCR 3.	

334-433 / SCARCS

Use any 8 bits No.3

```

FT DOMAIN 500 601 SRCR 4.
FT DOMAIN 619 875 SERINE PROTEASE.
FT ACT_SITE 619 830 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA; 97011 MW; 67D52272B CRC32;

Query Match 100.0%; Score 787; DB 1; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.39e-183;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 387 IRLAGKSGHEGRLEYYRGQMGTVCDGTELTNTYVVCROLGFKYKQASANHFESTG 446
334 IRLAGKSGHEGRLEYYRGQMGTVCDGTELTNTYVVCROLGFKYKQASANHFESTG 393
447 PIWLDVSCSGKETRFLOCRRQWGRHDCSHREDVSIACY 486
394 PIWLDVSCSGKETRFLOCRRQWGRHDCSHREDVSIACY 433

RESULT 2 STANDARD; PRT; 761 AA.
ID NETR_MOUSE
AC 008762;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin) (BRAIN-SPECIFIC SERINE
DE PROTEASE 3) (BSP-3).
GN PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GSCHEWED T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neurotrypsin, a novel multidomain serine protease expressed in the
RL nervous system."
RL Mol. Cell. Neurosci. 9:207-219(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSUBOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
RT cringle-like structure and three scavenger receptor cysteine-rich
RT motifs."
RN Blochem. Biophys. Res. Commun. 239:386-392(1997).
OC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
OC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
OC LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
CC AND AMYGALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).

```

```

DR EMBL: Y13192; CAA73646.1; -
DR EMBL: D89871; BAA23986.1; -
DR MGD; MG1:1100881; PRSS12.
DR PFAM; PF00530; SRCR; 3.
DR PFAM; PF00089; trypsin; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PROSITE; PS00420; SPERACT_RECEPTOR; 3.
KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 761
FT DOMAIN 85 157 NEUROTRYPsin.
FT DOMAIN 166 267 KRINGLE.
FT DOMAIN 273 373 SRCR 1.
FT DOMAIN 386 487 SRCR 2.
FT DOMAIN 505 761 SRCR 3.
FT DOMAIN 505 761 SERINE PROTEASE.
FT ACT_SITE 516 517 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 516 517 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT DISULFID 711 711 CHARGE RELAY SYSTEM.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA; 84118 MW; 3F3C4F35 CRC32;

Query Match 87.8%; Score 691; DB 1; Length 761;
Best Local Similarity 82.0%; Pred. No. 7.85e-157;
Matches 82; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 273 IRLAGKSGHEGRLEYYRGQMGTVCDGTELTNTYVACRLGFEYKQSSVNHFDGSR 332
334 IRLAGKSGHEGRLEYYRGQMGTVCDGTELTNTYVACRLGFEYKQSSVNHFDGSR 393
333 PIWLDVSCSGKEVFQICSRQWGRHDCSHREDVGLICY 372
394 PIWLDVSCSGKETRFLOCRRQWGRHDCSHREDVSIACY 433

RESULT 3 STANDARD; PRT; 453 AA.
ID MSRE_BOVIN
AC P21758;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUOIRA P.,
RA KRIGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
RT collagen-like coiled coils."
RN Nature 343:570-572(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
RT scavenger receptor type II."
RN Nature 343:570-572(1990).
CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING Atherosclerosis
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE
CC ENOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING

```

CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC -1- SUBUNIT: HOMOTRIMER.
CC CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
CC OF THE SAME GENE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X51689; CAA35987.1; -.
EMBL: X54183; CAA38108.1; -.
PIR: S08276; S08276.
PIR: S08278; S08278.
DR PROSITE; PS00420; SPERACT_RECEPTOR; 1.
DR PFMAM; PF00530; SRCR; 1.
DR PFMAM; PF01391; Collagen; 1.
KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
KM Heptad repeat pattern; Receptor; Alternative splicing.
FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 51 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 77 453 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 77 108 SPACER (PROBABLE).
FT DOMAIN 109 271 ALPHA-HELICAL COILED-COIL.
FT DOMAIN 272 343 COLLAGEN-LIKE.
FT DOMAIN 344 453 CYS-RICH.
FT CARBOHYD 82 82 POTENTIAL.
FT CARBOHYD 101 101 POTENTIAL.
FT CARBOHYD 142 142 POTENTIAL.
FT CARBOHYD 183 183 POTENTIAL.
FT CARBOHYD 220 220 POTENTIAL.
FT CARBOHYD 248 248 POTENTIAL.
FT CARBOHYD 266 266 POTENTIAL.
FT VARSPPLIC 348 349 OS -> PG (IN ISOFORM II).
FT VARSPPLIC 350 453 MISSING (IN ISOFORM II).
SQ SEQUENCE 453 AA; 50056 MW; 2CDEIEI CRG32;

Query Match 52.5%; Score 413; DB 1; Length 453;
Best Local Similarity 52.9%; Pred. No. 7,34e-81;
Matches 54; Conservative 19; Mismatches 24; Indels 5; Gaps 4;

```

352 VRLVGGSGPRHGGRVFIHFEGMGWGYTCDDRR-ELRGGLVYGSLSGLK-GVQSHKRATYGCK 409
      ::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
334 IRLVGKSKSHGRLRELVYRGWGTYCDGDGWL-NTYYVCQLDGFRKYGQSASNH-FEE 390
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db    410 GTGPIMLENEVCFGKSSIEECRIKMGVCRACSHDEDGVGTC 451
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy    391 STGPIMLDVSCSGKETRFLOCSRQRMRHDCSHREDVSITAC 432
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT          4
ID MSRE_HUMAN   STANDARD;             PRT;           451 AA.
AC P21757; P21759;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
DE LDL RECEPTOR I AND II).
GN MSRL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91067661.
RA MATSUOTO A., NAITO M., ITAKURA H., IKEMOTO S., ASAOKA H.,
RA HATAKAWA I., KANAMORI T., ABDURAZANI H., TAKAKU F., SUZUKI H.,
RA KOBARI Y., MIYAI T., TAKAHASHI K., COHEN E.H., WUDRO R.,

```

```

RA HOUSMAN D.E., KODAMA T.:  

RT "Human macrophage scavenger receptors: primary structure, expression,  

RI and localization in atherosclerotic lesions".  

RL Proc. Natl. Acad. Sci. U.S.A. 87:9133-9137(1990).  

CC -!- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC  

CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.  

CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE  

CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING  

CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).  

CC -!- SUBUNIT: HOMOTRIMER.  

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  

CC -!- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER  

CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING  

CC OF THE SAME GENE.  

-----  

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  

CC the European Bioinformatics Institute. There are no restrictions on its  

CC use by non-profit institutions as long as its content is in no way  

CC modified and this statement is not removed. Usage by and for commercial  

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  

CC or send an email to license@isb-sib.ch).  

-----  

CC EMBL; D90187; BAA14208.1; .  

CC DR EMBL; D90188; BAA14209.1; .  

CC DR PIR; A38415; A38415.  

CC DR PIR; B38415; B38415.  

CC DR MIM; 153622; -  

CC DR PROSITE; PS00420; SPERACT_RECEPTOR; 1.  

CC DR PFAM; PF00530; SRCR; 1.  

CC DR PFAM; PF01391; Collagen; 1.  

CC KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;  

CC HV Heptad repeat pattern: Receptor: Alternating splicing.  

CC FM DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).  

CC FT TRANSMEM 51 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  

CC FT DOMAIN 77 451 EXTRACELLULAR (POTENTIAL).  

CC FT DOMAIN 77 109 SPACER (PROBABLE).  

CC FT DOMAIN 110 272 ALPHA-HELICAL COILED-COIL.  

CC FT DOMAIN 273 341 COLLAGEN-LIKE.  

CC FT DOMAIN 342 451 CYS-RICH.  

CC FT CARBOHYD 82 82 POTENTIAL.  

CC FT CARBOHYD 102 102 POTENTIAL.  

CC FT CARBOHYD 143 143 POTENTIAL.  

CC FT CARBOHYD 184 184 POTENTIAL.  

CC FT CARBOHYD 221 221 POTENTIAL.  

CC FT CARBOHYD 249 249 POTENTIAL.  

CC FT CARBOHYD 267 267 POTENTIAL.  

CC FT VASPLPIC 345 358 TPPTKVLVVGSGP -> RPVALTHIRAGPS (IN  

CC ISOFORM II).  

CC VS VASPLPIC 359 451 MISSING (IN ISOFORM II).  

CC FT SEQUENCE 451 AA; 49762 MW; AB9PEEC7 CRC32.;  

Query Match 51.7%; Score 407; DB 1; Length 451;  

Best Local Similarity 51.5%; Pred. No. 2,93e-79;  

Matches 52; Conservative 18; Mismatches 28; Indels 3; Gaps 2;  

Db 350 VRIVGSGPHRGREYLIHLSGMGTICDDRWVRVRYGVCSRSLGR-VQVAHNKAHFGGG 408  

Cc :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:::  

Oy 334 IRLGGKGSHSGRGLRVYIRGQWGTVCDDGWELNTIYYCRQLGKYKQA--SANHEES 391  

Cc |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:::  

Db 409 TGPIWLNEVEFCGRESSIEECKIKOMQTGRACSHSDAGATC 449  

Cc |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:::  

Oy 392 TGPIMLDVSCSGKETRFLOCSRRONGRHDCSHDEDVSINC 432  

Cc |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:::  

RESULT 5  

ID MSRE_RABIT STANDARD; PRT; 454 AA.  

AC 005585;  

DT 01-OCT-1996 (Rel. 34, Created)  

DT 01-OCT-1996 (Rel. 34, Last sequence update)  

DT 01-OCT-1996 (Rel. 34, Last annotation update)  

DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED  

DE LDL RECEPTOR I AND II).

```

```

MSR1.
GN Oryctolagus cuniculus (Rabbit),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
RN Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RA [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93016877.
RA BICKEL P.E., FREEMAN M.W.;
RT "Rabbit aortic smooth muscle cells express inducible macrophage
RT scavenger receptor messenger RNA that is absent from endothelial
RL cells";
RN J. Clin. Invest. 90:1450-1457(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93131972.
RA DOI T., WADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
RA MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KAWABE Y.,
RA ITAHARA H., YAZAKI Y., MATSUMOTO A.;
RT "Charged collagen structure mediates the recognition of negatively
RT charged macromolecules by macrophage scavenger receptors.";
RN J. Biol. Chem. 268:2126-2133(1993).
-1- FUNCTION: MEMBRANE GLYCOPROTEIN; IMPLICATED IN THE PATHOLOGIC
CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
CC OF THE SAME GENE.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL, L11693; AAA31402.1; -.
DR EMBL, L11692; AAA31403.1; -.
DR EMBL, D13381; BAA02649.1; -.
DR PROSITE, PS00420; SPERACT_RECEPTOR, 1.
DR PFMW, PF00530; SRCR, 1.
KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
KW Heptad repeat pattern; Receptor; Alternative splicing.
FT DOMAIN 1 50
FT TRANSMEM 51 73
DR DOMAIN 74 454
DR DOMAIN 74 454
FT DOMAIN 110 272
FT DOMAIN 273 344
FT DOMAIN 345 454
FT CARBOHYD 82 82
FT CARBOHYD 102 102
FT CARBOHYD 143 143
FT CARBOHYD 184 184
FT CARBOHYD 221 221
FT CARBOHYD 249 249
FT CARBOHYD 267 267
FT CARBOHYD 354 354
FT VARSPPLIC 355 454
FT CONFLICT 106 106
SO SEQUENCE 454 AA; 49745 MW; 5D780348 CRC32;
Query Match 51.5%; Score 405; DB 1; Length 454;
Best Local Similarity 50.5%; Pred. No. 1,00e-78;
Matches 51: Conservative 22; Mismatches 25; Indels 3; Gaps 3;

```

[illegible]

DR	EMBL:	L04274:	AAA39747.1:	ALT_INT.
DR	EMBL:	L04275:	AAA39748.1:	ALT_INT.
DR	EMBL:	D13382:	BAA02650.1:	-
DR	EMBL:	M59445:	AAA37464.1:	-
DR	EMBL:	M59446:	AAA37465.1:	-
DR	EMBL:	U13873:	AAC13774.1:	-
DR	PIR:	A38260:	A38260.	-
DR	MCD:	MGI:98257:	SCVR.	
DR	PROSITE:	PS00420:	SPRAC1_RECEPTOR; 1.	
DR	PFAM:	PF00530:	SRCR; 1.	
DR	PFAM:	PF01391:	Collagen; 1.	
KW	Transmembrane:		Glycoprotein; Endocytosis; Coiled coil; LDL; Heptad repeat pattern; Receptor; Alternative splicing.	
FT	DOMAIN	1	55	CYTOLASMIC (POTENTIAL).
FT	TRANSMEM	56	78	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN	79	458	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	79	114	SPACER (PROBABLE).
FT	DOMAIN	114	276	ALPHA-HELICAL COILED-COIL.
FT	DOMAIN	277	350	COLLAGEN-LIKE.
FT	DOMAIN	351	458	CYS-RICH.
FT	CARBOHYD	94	94	POTENTIAL.
FT	CARBOHYD	107	107	POTENTIAL.
FT	CARBOHYD	147	147	POTENTIAL.
FT	CARBOHYD	188	188	POTENTIAL.
FT	CARBOHYD	253	253	POTENTIAL.
FT	CARBOHYD	271	271	POTENTIAL.
FT	VASAPPLIC	352	354	TPL. -> RSV (IN ISOFORM II).
FT	VASAPPLIC	355	458	MISSING (IN ISOFORM II).
SO	SEQUENCE	458 AA:	50130 MW:	F28A456E CRC32:

Query Match	49.98;	Score 393;	DB 1;	Length 458;
Best Local Similarity	47.08;	Pred. No. 1.56e-75;		
Matches	47;	Conservative	18;	Mismatches 34;
			Indels	1;
			Gaps	1

[illegible]

RESULT	7	
WC11_BOVIN		STANDARD;
TR		PRT; 1436 AA

DE 01-APR-1993 (Rel. 25, Last sequence update)
OS 01-APR-1993 (Rel. 25, Last sequence update)
OC 01-FEB-1984 (Rel. 28, Last annotation update)
DE ANTIGEN WCI.1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD.
RX MEDLINE; 93056489.
RA WINGGAARD P.L.J., METZELAR M.J., MACHUGH N.D., MORRISON W.I.,
RA CLEVERS H.C.;
RT "Molecular characterization of the WCI antigen expressed specifically
RT on bovine CD4-CD8-gamma delta T lymphocytes.";
RL J. Immunol. 149:3273-3277(1992)

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sb.ch).

```
DR EMBL; X63723; CAA55255.1; -.
DR PIR: S19913; S19913.
DR PIR: A46495; A46495.
DR PROSITE; P500420; SEPRACT_RECEPTOR; 4.
DR FRAM; PF00350; SRCK; 11.
DR Antigen; Repeat.
KW
SEQUENCE 1436 AA; 154196 MW; D61545AD7 CRC32
```

Query Match	49.7%;	Score 391;	DB 1;	Length 1436;
Best Local Similarity	47.5%;	Pred. No. 5.32e-75;		
Matches	48;	Conservative	26;	Mismatches 24;
			Indels	3;
			Gaps	3

```

Db 1RLVDGGPCCGAREIILDGSGMGTICDDW-DLIDARVYECGLGCGALNATSAHGAG 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 334 IRLANGKSHHGRLVYVYRGCMGTVCDDGWTELN-TVYCRQLGFKYG-RQASNHHEES 39
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 990 SGPTIMDLDLNCTGKESHVWRPSPSGMGRHDCRHEDPADGAVLC 1030
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 392 TGPTIMLDVSCSGKETRFLQCSRKQMGHHDCSHEDDSIAC 432

```

RESULT	8
CD6_HUMAN	STANDARD;
	PRT;
	468 AA

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE T-CELL DIFFERENTIATION ANTIGEN CD6 PRECURSOR (T12) (TP120).
GN CD6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92013796.
RA ARUFFO A., MELNICK M.B., LINSLEY P.S., SEED B.;
RT "the lymphocyte glycoprotein CD6 contains a repeated domain structure
RT characteristic of a new family of cell surface and secreted
RT proteins.";
RT J. Exp. Med. 174:949-952(1991).
RL
RL FUNCTION: INVOLVED IN CELL ADHESION. BINDS TO CD166.
CC
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC
CC -1- SIMILARITY: CONTAINS 3 SRC DOMAINS.
CC
CC -1- DATABASE: NAME-PROT; NOTE-CD guide CD6 entry;
CC WWW=<http://www.ncbi.nlm.nih.gov/prov/cd/cd6.htm>.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; X60992; CAA43306.1;	-
DR	PIR; S26741; S26741.	
DR	MIM; 186720;	-
DR	PROSITE; PS00420; SPBRCT_RECEPTOR; 1.	
DR	Pfam; PF00530; SRCF; 3.	
Kw	Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal.	
FT	SIGNAL	1 ?
FT	CHAIN	? 468
FT	DOMAIN	? 402
FT	TRANSMEM	403 423
FT	DOMAIN	424 468
FT	CARBOHYD	28 28
FT	CARBOHYD	49 49
FT	CARBOHYD	112 112
FT	CARBOHYD	118 118
FT	CARBOHYD	229 229
FT	CARBOHYD	339 339
FT	CARBOHYD	345 345
FT	CARBOHYD	368 368

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Catartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-DUODENAL MUCOSA;
 RX MEDLINE: 94327548.
 RA MATSUHARA M., ICHINOSE M., YAHAGI N., KAKEI N., TSUKADA S.,
 MIKI K., KUROKAWA K., TASHIRO K., SHIOKAWA K., SHINOMIYA K.,
 UMEYAMA H., INOUE H., TAKAHASHI T., TAKAHASHI K.;
 RL "Structural characterization of porcine enteropeptidase";
 J Biol Chem. 269:19976-19982(1994).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN
 CC -1- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
 (HEAVY) CHAIN, AND A MINI CHAIN.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 NAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D30799; BAA06459.1; -.
 CC HSSP: P00763; LDPO.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 PFAM: PF00057; Ldl_recept_a; 2.
 PFAM: PF00089; trypsin; 1.
 PFAM: PF00431; CUB; 2.
 PFAM: PF00530; SRCR; 1.
 PFAM: PF00629; MAM; 1.
 PFAM: PF01380; SEA; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 52 117
 FT CHAIN 118 799
 FT CHAIN 800 1034
 FT TRANSMEM 19 47
 FT DOMAIN 197 238
 FT DOMAIN 240 349
 FT DOMAIN 357 519
 FT DOMAIN 539 649
 FT DOMAIN 656 694
 FT DOMAIN 693 786
 FT ACT_SITE 840 840
 FT ACT_SITE 891 891
 FT ACT_SITE 986 986
 FT LIPID 2 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 BY SIMILARITY.

FT DISULFID 658 670 BY SIMILARITY.
 FT DISULFID 665 683 BY SIMILARITY.
 FT DISULFID 677 692 BY SIMILARITY.
 FT DISULFID 787 911 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 825 841 BY SIMILARITY.
 FT DISULFID 925 992 BY SIMILARITY.
 FT DISULFID 956 971 BY SIMILARITY.
 FT DISULFID 982 1010 BY SIMILARITY.
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 283 283
 FT CARBOHYD 343 343
 FT CARBOHYD 350 350
 FT CARBOHYD 403 403
 FT CARBOHYD 455 455
 FT CARBOHYD 485 485
 FT CARBOHYD 518 518
 FT CARBOHYD 549 549
 FT CARBOHYD 645 645
 FT CARBOHYD 697 697
 FT CARBOHYD 701 701
 FT CARBOHYD 721 721
 FT CARBOHYD 740 740
 FT CARBOHYD 761 761
 FT CARBOHYD 804 804
 FT CARBOHYD 863 863
 FT CARBOHYD 902 902
 FT CARBOHYD 964 964
 SQ SEQUENCE 1034 AA; 114776 MW; 24386471 CRC32;
 Query Match 13.5%; Score 106; DB 1; Length 1034;
 Best Local Similarity 27.4%; Pred. No. 1.75e-05;
 Matches 17; Conservative 17; Mismatches 27; Indels 1; Gaps 1;
 Db 693 VREFLNGTANSGVIOFRIQSIHTACAEWNTQTSDDVCCOLLIGTGN-SMPEFFSSGG 751
 QY 334 IRLAGKSGSHGRLVYRQGWTVCDGWTDLNTYVVCNQLGPKYKQKSAHFEESTG 393
 Db 752 PF 753
 QY 394 PI 395
 RESULT 12
 ID CD5_HUMAN STANDARD; PRT; 495 AA.
 AC P06127;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCYTE GLYCOPROTEIN
 DE TI/LEU-1) (LYMPHOCYTE ANTIGEN CD5).
 GN CD5 OR LEU1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87014786.
 RA JONES N.H., CLABBY M.L., DIALYNAS D.P., HUAG H.-J.S.,
 RA HERZENBERG L.A., STROWINGER J.L.;
 RT "Isolation of complementary DNA clones encoding the human lymphocyte
 glycoprotein TI/Leu-1";
 RL Nature 323:346-349(1986).
 RN [2]
 RP INTERACTION WITH CD72/LYB-2.
 RX MEDLINE: 91270374.
 RA VAN DE VELDE H., VON HOEGEN I., LIO W., PARNES J.R., THIELEMANS K.;
 RT "The B-cell surface protein CD72/Lyb-2 is the ligand for CD5.";
 RL Nature 351:662-665(1991).
 CC -1- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL
 CC PROLIFERATION. CD5 INTERACTS WITH CD72/LYB-2.

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 SRCR DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd5.htm"
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04391; CAA27979.1; -.
DR PIR: A26396; A26396.
DR MIM: 153340; -.
DR PIRAM: PF00530; SRCR; 3.
KW Signal; Transmembrane; Glycoprotein; T-cell.
KV SIGNAL 1 24
FT TMHAIN 25 495 T-CELL SURFACE GLYCOPROTEIN CD5.
FT DOMAIN 25 372 EXTRACELLULAR (POTENTIAL).
FT TRASNMM 373 402 POTENTIAL.
FT DOMAIN 403 495 CYTOPLASMIC (POTENTIAL).
FT TMHAIN 34 133 SRCR 1.
FT DOMAIN 275 382 SRCR 2.
FT DISULFD 44 107 POTENTIAL.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 241 241 POTENTIAL.
FT SEQUENCE 495 AA; 54625 MW; 8507C34C CRC32;

Query Match 13.3%; Score 105; DB 1; Length 495;
Best Local Similarity 37.9%; Pred. No. 2,72e-05;
Matches 11; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

Db 36 RLTRNSKQGOLEYLYKXDMHVCOSQW 64
   ||::: ||::|::|::|::|::|::|::|::|::|
Qy 335 RLAGGKGSHEGRLEYYRQMGTVCDGW 363

RESULT 13
ID TMG2_HUMAN STANDARD: PRT; 492 AA.
AC O15393;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).
GN TMPSR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
[11]
CC SEQUENCE FROM N.A.
CC MEDLINE: 97468144.
CC RA PAOLONI-GIACOBINO A., CHEN H., PETTSCH M.C., ROSSIER C.,
CC ANTONARAKIS S.E.;
CC "Cloning of the TMPSR2 gene, which encodes a novel serine protease
CC with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3";
CC Genomics 44:309-320(1997).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND
CC WEAKLY IN SEVERAL OTHER TISSUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRIPSTIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: U75329; AAC51784.1;
DR HSSP; P00763; IDPO.
DR MIM; 602060;
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PROSITE; PS01209; LDLRA_2; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
KW Hydroxylase; Serine protease; Transmembrane; Signal anchor.
FT DOMAIN 1 84 CYTOSOL PLASMIC (POTENTIAL).
FT TRANSMEM 85 105 POTENTIAL.
FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 111 150 LDL-RECEPTOR CLASS A.
FT DOMAIN 151 243 SRCR.
FT DOMAIN 256 492 CATALYTIC.
FT ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 POTENTIAL.
FT CARBOHYD 249 249 POTENTIAL.
SQ SEQUENCE 492 AA; 53847 MW; 19BC1F67 CRC32;

Query Match 12.7%; Score 100; DB 1; Length 492;
Matches Similarity 40.7%; Pred. No. 2,40e+04;
Match 11; Conservative 4; Mismatches 12; Indels 0; Gaps

DB 165 RKSHPVCCDDMMENYGRACPRDGMGX 191
QY 352 RGMCTGVCCDDMTLNTYVVCROLGFR 378

RESULT 14
ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6; TISSUE=DUDENUM;
RA YUON X., LU D., RUBIN D.C., PUNG C.Y.M., SADDLER J.E.;
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPsin AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPsinOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLY).
CC -1- PMV: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

```



```
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS00740; MAM; 1.
DR PROSITE; PS50060; MAM; 2; 1.
DR PROSITE; PS01209; LDLRA; 1; 2.
DR PROSITE; PS50068; LDLRA; 2; 2.
DR PFAM; PF00057; ldl_recept_a; 2.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00530; SRCR; 1.
DR PFAM; PF00629; MAM; 1.
DR PFAM; PF01390; SEA; 1.
KW Serine protease; Zymogen; Transmembrane; Repeat;
  Signal-anchor; Glycoprotein; Myristate; Hydrolase;
  NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 1 784
FT CHAIN 19 47
FT TRANSMEM 182 223
FT DOMAIN 225 334
FT DOMAIN 342 504
FT DOMAIN 524 634
FT DOMAIN 641 679
FT ACT_SITE 678 771
FT ACT_SITE 825 825
FT ACT_SITE 876 876
FT ACT_SITE 971 971
FT LIPID 2 2
FT DISULFID 184 197
FT DISULFID 191 210
FT DISULFID 204 221
FT DISULFID 643 655
FT DISULFID 650 668
FT DISULFID 662 677
FT DISULFID 772 896
FT DISULFID 810 826
FT DISULFID 910 977
FT DISULFID 941 956
FT DISULFID 967 995
FT CARBOHYD 116 116
FT CARBOHYD 147 147
FT CARBOHYD 179 179
FT CARBOHYD 328 328
FT CARBOHYD 335 335
FT CARBOHYD 388 388
FT CARBOHYD 440 440
FT CARBOHYD 470 470
FT CARBOHYD 503 503
FT CARBOHYD 534 534
FT CARBOHYD 630 630
FT CARBOHYD 682 682
FT CARBOHYD 706 706
FT CARBOHYD 725 725
FT CARBOHYD 848 848
FT CARBOHYD 887 887
FT CARBOHYD 909 909
FT CARBOHYD 949 949
SQ SEQUENCE 1019 AA; 112923 MW; 0E641C53 CRC32;

Query Match 12.6%; Score 99; DB 1; Length 1019;
Best Local Similarity 31.3%; Pred. No. 3.68e-04;
Matches 20; Conservative 14; Mismatches 25; Indels 5; Gaps 4;

Db 678 VREFNGTNNNG-L-VREFQSIWHTACAEWNVTQISNDVCOILGLGSGNSKPI-FSTD 734
QY 334 IRLAGKSGSHGRLVYRGQ--WGTVCDGDWTELTNYVVCROIGFKYKQASANHFEE 391
Db 735 GGPF 738
QY 392 TGPI 395
```

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

h_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Mon Mar 13 10:30:49 2000;      MasPar time 16.65 Seconds
Tabular output not generated.              416.315 Million cell updates/sec
```

```

Title: >US-09-147-947-6
Description: (334-433) from US09147947A.ppt (5 of 6)
Perfect Score: 787
Sequence: 1 IRLGGGSHGRLFVYTRG.....RQQRHDCRHEDVSTACY 100

```

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

```
post-processing: Minimum Match 08
                  Listing first 45 summaries
```

```
Database:
  sprenb12
    1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
    5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
    9:sp.phage 10:sp.plant 11:sp.podent 12:sp.unclassified
    13:sp.vertebrate 14:sp.virus
```

Statistics: Mean 37.816; Variance 60.520; scale 0.625

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	432	54.9	2043	5	065943	SRCR DOMAIN, MEMBRANE	1.87e-79
2	419	53.2	918	13	029098	PEMA-SRR PROTEIN PREC	3.1e-76
3	412	52.4	1116	4	007898	M130 ANTIGEN, EXTRACT	1.72e-74
4	412	52.4	1149	4	007901	M130 ANTIGEN, CYTOLPLAS	1.72e-74
5	412	52.4	1151	4	007899	M130 ANTIGEN, CYTOLPLAS	1.72e-74
6	412	52.4	1156	4	007900	M130 ANTIGEN, CYTOLPLAS	1.72e-74
7	412	52.4	2153	5	097375	SCAVENGER-RECEPTOR CVS	1.46e-65
8	383	48.7	369	6	029110	SCAVENGER-RECEPTOR PRO	2.46e-57
9	383	48.7	774	4	09Y4K0	LYSYL OXIDASE-RELATED	1.34e-57
10	380	48.3	546	6	029111	SCAVENGER-RECEPTOR PRO	1.34e-56
11	374	47.5	600	6	028910	MUCIN (FRAGMENT)'. HENSLIN.	3.99e-65
12	369	46.9	1594	6	095218	MEMBRANE PROTEIN SCAVE	6.69e-64
13	365	46.4	462	6	028881	PANCREAS CANCER-ASSOCI	1.12e-62
14	364	46.3	578	11	P70117	MAC-2 BINDING PROTEIN	1.97e-62
15	363	46.1	585	4	008380	LYSYL OXIDASE RELATED	3.45e-62
16	362	46.0	895	13	09W6N1	DMBT1/6EB.1 PROTEIN PR	3.45e-62
17	362	46.0	1785	4	09Y4V9	DMBT1 PROTEIN.	3.45e-62
18	362	46.0	1785	4	09Y211	DMBT1 PROTEIN.	3.45e-62
19	361	45.9	374	11	070513	MAMA.	6.06e-62
20	360	45.7	1036	5	09Y378	SCAVENGER RECEPTOR CVS	1.05e-61

21	353	44.9	352	11	035301	SP-ALPHA.	5.42e-60
22	353	44.9	352	11	035301	SP-ALPHA.	5.42e-60
23	352	44.7	518	11	060754	BACTERIA BINDING MACRO	9.50e-60
24	351	44.6	1290	11	062827	ENTERIN.	1.66e-59
25	350	44.5	577	11	007797	PEPTIDYLPROLYL ISOMERA	2.92e-59
26	350	44.5	2083	11	060997	CAR-DUCIN PRECURSOR (2.92e-59
27	347	44.1	520	4	09553	MACROPHAGE RECEPTOR.	1.57e-58
28	346	44.0	504	6	028908	MUCIN (FRAGMENT).	2.74e-58
29	343	43.6	347	4	043865	SP ALPHA.	1.47e-57
30	339	43.1	483	11	09W0B9	MACROPHAGE RECEPTOR MA	1.37e-56
31	339	43.1	574	11	035649	CYCLOPHILIN C-ASSOCIAT	1.37e-56
32	332	42.2	480	6	029112	SCAVENGER-RECEPTOR PRO	6.81e-55
33	331	42.1	804	6	029113	SCAVENGER-RECEPTOR PRO	1.19e-54
34	326	41.4	356	6	097682	UNKNOWN MRNA, PARTIAL	1.92e-53
35	326	41.4	754	11	092175	LYSYL OXIDASE-RELATED	1.92e-53
36	320	40.7	437	6	029109	SCAVENGER-RECEPTOR PRO	5.37e-52
37	318	40.4	127	6	095316	M130 ANTIGEN (FRAGMENT	1.63e-51
38	308	39.1	528	5	097379	SCAVENGER RECEPTOR CTS	4.10e-49
39	308	39.1	531	5	017064	SP85.	4.10e-49
40	300	38.1	822	13	09YHC1	LYSYL OXIDASE HOMOLOG	3.36e-47
41	299	38.0	638	4	09Y5V8	LYSYL OXIDASE-Like PRO	5.83e-47
42	293	37.2	592	4	09Y4K7	CD5E.	1.57e-45
43	293	37.2	595	4	09Y4K9	CD6C.	1.57e-45
44	293	37.2	601	4	09Y4K8	CD6D.	1.57e-45
45	293	37.2	635	4	09Y4L0	CD6E.	1.57e-45

ALIGNMENTS

RESULT	1		PRELIMINARY;	PRT; 2043 AA.	<i>dog hole</i>
ID	O96943				
AC	O96943:				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1998	(TREMBLrel. 10, Last sequence update)			
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)			
DE	SRCR DOMAIN, MEMBRANE FORM 2.				
GN	SRCRM2.				
OS	Gedidia cydonium (Sponge).				
OC	Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;				
CC	Astrophorida; Geodidae; Geodia.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 98369060.				
RA	BUMBACH B., PANCER Z., DIEHL-SEIFERT B., STEFFEN R., MOENNERN J.,				
RA	MUELLER I., MUELLER W.E.G.:				
RT	"The putative sponge aggregation receptor. Isolation and				
RT	characterization of a molecule composed of scavenger receptor				
RT	cysteine-rich domains and short consensus repeats."				
RL	J. Cell Sci. 111:2635-2644(1998).				
DR	EMBL; Y14953; CAA75175.1; .				
DR	HSSP; P10998; IVVC.				
SO	PROSITE; PS00420; SPERACT_RECEPTOR; 8.				
SEQUENCE	2043 AA; 220896 MW; 4CE19401 CRC32;				
Query Match		54.9%; Score 432; DB 5; Length 2043;			
Best Local Similarity	51.0%; Pred. No. 1.87e-79;				
Matches	51; Conservative	21; Mismatches 27; Indels 1; Gaps 1			
Dd	936 IRLVGSGSPHGGRVEIYQGVGVTCDSDSWGCPADAVYCROLGYANASRATVRAEFGRG	995			
Oy	334 IRLAGKSKSHSRLEVVIRGOMGVICDDGWELMTYYVCRQLGRKYRKQAASA-NHFEEST	392			
Dd	996 GEIMLDNVACTGTGFENSIDECRSNGMGDHNCGRDAGAVC	1035			
Oy	393 GPWILDVSCSGKETRFLOCSRGMWRGHDCSHREDVSIAIC	432			
RESULT	2		PRELIMINARY;	PRT; 918 AA.	
ID	O92098				
AC	O92098:				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)			

RN	[1]	SEQUENCE FROM N.A.
RP		MEDLINE: 93380506.
RX		LAW S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,
RA		MASON D.Y.;
RT		"A new macrophage differentiation antigen which is a member of the
RL		scavenger receptor superfamily.";
DR		Eur. J. Immunol. 23:2320-2325(1993).
EMBL:		J22899; CA80542.1; -.
PFAM:		PF00530; SRCR_9.
PRINTS:		PRO0258; SPERACTRCPTP.
KW		Antigen; Signal.
FT	SIGNAL	1 40 POTENTIAL.
CHAIN	41 1151	M130 ANTIGEN, CYTOPLASMIC VARIANT 1.
SO	SEQUENCE	1151 AA; 124820 MW; A7ZEDZF CRC32;
Query Match	52.4%;	Score 412; DB 4; Length 1151;
Best Local Similarity	48.5%;	Pred. No. 1.72e-74;
Matches	49; Conservative	25; Mismatches 24; Indels 3; Gaps 3
Db	714 LRLVNGGRCAGRVETIYHEGSNGITCDSDS-DLSDAHVCROLGGEAINTSGAHFEG	772
QY	334 IRLAGKGSHGRLELVYRGWGVTCCDDGWTEL-NTYYVCRQLTFKYKQASAN-HFEES	391
OY	392 TGPILWDVKSCGKETRFLOCSRRQMGHDCSHREDYSIAC	432
RESULT	6 PRELIMINARY;	PRT: 1156 AA.
ID	007900.	
AC	007900.	
DT	01-NOV-1996 (TREMBLrel. 01. Created)	
DT	01-NOV-1996 (TREMBLrel. 01. Last sequence update)	
DE	01-NOV-1999 (TREMBLrel. 12. Last annotation update)	
DE	M130 ANTIGEN, CYTOPLASMIC VARIANT 2 PRECURSOR.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
[1]		
RP	SEQUENCE FROM N.A.	
RX	MEDLINE: 93380506.	
RA	LAW S.A., MICKLEK K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,	
RA	MASON D.Y.;	
RT	"A new macrophage differentiation antigen which is a member of the	
RT	scavenger receptor superfamily.";	
RL	Eur. J. Immunol. 23:2320-2325(1993).	
EMBL:	J22870; CA80543.1; -.	
PFAM:	PF00530; SRCR_9.	
PRINTS:	PRO0258; SPERACTRCPTP.	
KW	Antigen; Signal.	
FT	SIGNAL	1 40 POTENTIAL.
CHAIN	41 1156	M130 ANTIGEN, CYTOPLASMIC VARIANT 2.
SO	SEQUENCE	1156 AA; 125352 MW; 287A0780 CRC32;
Query Match	52.4%;	Score 412; DB 4; Length 1156;
Best Local Similarity	48.5%;	Pred. No. 1.72e-74;
Matches	49; Conservative	25; Mismatches 24; Indels 3; Gaps 3
Db	714 LRLVNGGRCAGRVETIYHEGSNGITCDSDS-DLSDAHVCROLGGEAINTSGAHFEG	772
QY	334 IRLAGKGSHGRLELVYRGWGVTCCDDGWTEL-NTYYVCRQLTFKYKQASAN-HFEES	391
OY	392 TGPILWDVKSCGKETRFLOCSRRQMGHDCSHREDYSIAC	432
RESULT	7 PRELIMINARY;	PRT: 2153 AA.
ID	097375.	
AC	097375.	
DT	01-MAY-1999 (TREMBLrel. 10. Created)	
DT	01-MAY-1999 (TREMBLrel. 10. Last sequence update)	

	01-NOV-1999 (FEBMRLrel_12, Last annotation update)
DE	SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12 PRECURSOR.
CN	SRRR12.
OS	Strongylocentrotus purpuratus (Purple sea urchin).
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
CC	Euechinozoa; Echinacea; Echinoida; Strongylocentroididae;
NC	Strongylocentrotus.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	PANCER Z., RAST J., DAVIDSON E.H.;
RT	"Evolution of the Immune System: Transcription Factors and Downstream
FT	Genes of the Mammalian Immune System Expressed in Sea Urchin
RL	Coelomocytes.";
DR	Immunogenetics 0:0-0(1999).
EML:	AF064259; AAD08654.1
PROSITE:	PS00420: SPERRCT_RECEPTOR; 15.
KW	Signal; Receptor.
FT	SIGNAL 1 16 POTENTIAL.
CHAIN	17 2153 SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN TYPE 12.
SEQUENCE	2153 AA; 226557 MW; CA3BA8A9 CRC32;
Query Match	52.4%; Score 412; DB 5; Length 2153;
Best Local Similarity	48.0%; Pred. No.1,72e-74;
Matches	48; Conservative 25; Mismatches 26; Indels 1; Gaps 1;
Dd	684 MRLVGGNSRGRLGISINNQGTVCDDSDMDATVVCRCQLGSSVASPTSAHFGOGS 743
Yy	334 IRLAGKSGHEGRLEVYYRGCGVTCCDGDTELTNYVVCROLGFRKY-KOASAHFEEST 392
Dd	744 GTIIMLDVSCAGNSMLDGCRSLGVGHNCNAHADDAVYC 783
Yy	333 GPIIMLDVSCGKETRFLOCSRROGHHDCSHRDVSIAIC 432

```

RESULT      8      PRELIMINARY;      PRT;      369 AA.
ID Q29110
AC Q29110;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SCAVENGER-RECEPTOR PROTEIN (FRAGMENT).
GN WCI.
OS scrofa (Pig).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC (1)
RP SEQUENCE FROM N.A.
RN STRAIN=SLAB/B;
RX MEDLINE; 97364683.
RA KAHAN J.H.C., NAYEEM N., BINNS R.M., CHAIN B.M.;
RT "Mechanisms for variability in a member of the scavenger-receptor
RT cysteine-rich superfamily.";
RL Immunogenetics 46:276-282(1997).
DR EMBL; X99333; CAA67707.1; -.
DR PROSITE; PS00420; SPERACT_RECEPTOR; 2.
DR Pfam; PF00530; SRCR; 3.
FT NON_TER 1
FT NON_TER 369
FT NON_TER 369
SQ SEQUENCE 369 AA; 39710 MW; C669B4B CRC32;

Query Match 48.7%; Score 383; DB 6; Length 369;
Best Local Similarity 48.0%; Pred. No. 2,46e-67;
Matches 49; Conservative 23; Mismatches 25; Indels 5; Gaps 5;

Db 35 LRLVYG-GSYSGSEVELIHOGSWGTCDSM-DLDDAHVVCROLDGCKTSLALGSAHFGA 92
::: |||::: |||::: ||| ||| |||::: ||| ||| |||::: |||::: ||
Qy 334 IRLAGKGSH-EGRLVEYRGQMTVDGDWTELT-IYVVCRLDQ-FKRYKQASAHFEE 350
::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: ||

Db 93 GSGRWLDDVNCITKESHLMQCPBGRGQGNCRKADGAYIC 134
::: ||| |||::: |||::: |||::: |||::: |||::: |||::: |||::: ||
Qy 391 STGIWLDVSCSGKETRFLOCSRRQGRDSCSHREDVSIAC 432

```



```
RESULT 13
ID Q2881 PRELIMINARY; PRT; 462 AA.
AC Q2881;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MEMBRANE PROTEIN SCAVENGER RECEPTOR HOMOLOG (FRAGMENT).
GN T19.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
[1]
SEQUENCE FROM N.A.
MEDLINE: 95169648.
O'KEEFE M.A., METCALFE S.A., GLEN M.D., BOWDEN T., MCINNES S.,
KIMPTON W.G., CAHILL R.N., HEIN W.R., WALKER I.D.,
"lymph node homing cells biologically enriched for gamma delta T cells
express multiple genes from the T19 repertoire."
RT Int. Immunol. 6:1687-1697(1994).
RL EMBL: S76313; AAB33543.1; -.
DR EMBL: S76311; AAB33543.1; JOINED.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
DR PFAM: PF00530; SRCR; 3.
FT NON_TER 1
SQ SEQUENCE 462 AA; 49776 MW; 63EADB2 CRC32;

Query Match 46.4%; Score 365; DB 6; Length 462;
Best Local Similarity 43.6%; Pred. No. 6,38e-63;
Matches 44; Conservative 26; Mismatches 28; Indels 3; Gaps 3;

Db 4 LRLVGGCPAGREILIDQXSGXICDDG-DLDDAYVYCRGCGPALNATGSAFGAG 62
QY 334 IRLAGGSGHGRLEYYRGQWGTVCDDGTELTN-TYVCRQLGFKYG-KQASNHEES 391
OY 392 TGPILDDVSCSGKETRFLOCSRQWGRHDCSHREDVSIAC 432

RESULT 14
ID P70117 PRELIMINARY; PRT; 578 AA.
AC P70117;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PANCREAS CANCER-ASSOCIATED PROTEIN 4.
GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
SEQUENCE FROM N.A.
RA SCHAEFFERT C., POUR P.M., MACDONALD R.G., CHANEY W.G.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U73375; AAB18745.1; -.
DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
DR PFAM: PF00530; SRCR; 1.
DR PRINTS: PR00258; SPERACTRCPTR.
SQ SEQUENCE 578 AA; 64398 MW; 002E8BD2 CRC32;

Query Match 46.3%; Score 364; DB 11; Length 578;
Best Local Similarity 43.0%; Pred. No. 1.12e-62;
Matches 43; Conservative 22; Mismatches 34; Indels 1; Gaps 1;

Db 24 MRLVNGASANEGRVEIFYRGQWGTVCNDMLNILDANYVCALGYNATQALGRAAFGGR 83
QY 334 IRLAGGSGHGRLEYYRGQWGTVCDDGTELTN-TYVCRQLGFKYG-KQASNHEEST 392
OY 392 TGPILDDVSCSGKETRFLOCSRQWGRHDCSHREDVSIAC 432

Db 84 GPVVLDEVECTGTEPSLANCSLSGLKSRGCHENDAGVYC 123
QY 393 GPILDDVSCSGKETRFLOCSRQWGRHDCSHREDVSIAC 432
```

```
RESULT 15
ID Q08380 PRELIMINARY; PRT; 585 AA.
AC Q08380;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE MAC-2 BINDING PROTEIN PRECURSOR.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE: 93300818.
RX KOTHS K., TAYLOR E., HALENECK R., CASIPIT C., WANG A.;
RT "Cloning and characterization of a human Mac-2-binding protein, a new
RT member of the superfamily defined by the macrophage scavenger receptor
RT cysteine-rich domain."
RL J. Biol. Chem. 268:14245-14249(1993).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE: 94308070.
RA ULLRICH A., SURES I., D'EGIDO M., JALIAL B., POWELL T.J., HERBST R.,
RA DREPS A., AZAM M., ROBINSTEIN M., NATOLI C.;
RT "The secreted tumor-associated antigen 90K is a potent immune
RT stimulator."
RL J. Biol. Chem. 269:18401-18407(1994).
DR EMBL: L13210; AAA6193.1; -.
DR EMBL: X79089; CAA5699.1; -.
DR PFAM: PF00530; SRCR; 1.
DR PRINTS: PR00258; SPERACTRCPTR.
KW signal.
FT SIGNAL 1
FT CHAIN 19 POTENTIAL.
SQ SEQUENCE 585 AA; 65330 MW; AAER9E32 CRC32;

Query Match 46.1%; Score 363; DB 4; Length 585;
Best Local Similarity 43.0%; Pred. No. 1.97e-62;
Matches 43; Conservative 25; Mismatches 31; Indels 1; Gaps 1;

Db 24 MRLDGAINGRVEIFYRGQWGTVCNDMLNLTDSAVYCRALGFENATQALGRAAFGGS 83
QY 334 IRLAGGSGHGRLEYYRGQWGTVCDDGTELTN-TYVCRQLGFKYG-KQASNHEEST 392
OY 393 GPILDDVSCSGKETRFLOCSRQWGRHDCSHREDVSIAC 432

Db 84 GPVVLDEVECTGTEPSLANCSLSGLKSRGCHENDAGVYC 123
QY 393 GPILDDVSCSGKETRFLOCSRQWGRHDCSHREDVSIAC 432
```

Search completed: Mon Mar 13 10:31:07 2000
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

(TM)

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 585 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA: 65330 MW: 1874129 CN;

Query Match 51.7%; Score 391; DB 2; Length 585;
Best Local Similarity 52.0%; Pred. No. 6,80e-30;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

Db 24 MRLADGATNCRGVEIFYRGWGCTGCDNLMDLTDASVVCRAFGFNATQALGRAAFGGS 83
QY 447 VRLMDGKNKKEGVVEFYINGWGCTICDDGWTDKDAVICRQLGYGPARARIMAYFGGK 506
DB 84 GPIMLDEVCTGTEASLADCKSLGWLKSNCRHREDAGVC 123
OY 507 GPIHVDNVACTGNERSLADCIKQDIGHNCRHSEDAGVIC 546

2
US-08-477-674-10 STANDARD; PRT: 585 AA.

Sequence 10, Application US/08477674
Patent No. 5644035
GENERAL INFORMATION:
APPLICANT: Koths, Kirston E.
APPLICANT: Halenbeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,674
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 585 amino acids
TYPE: amino acid

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 585 AA: 65330 MW: 1874129 CN;

Query Match 51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 6,80e-30;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

Db 24 MRLADGATNCRGVEIFYRGWGCTGCDNLMDLTDASVVCRAFGFNATQALGRAAFGGS 83
QY 447 VRLMDGKNKKEGVVEFYINGWGCTICDDGWTDKDAVICRQLGYGPARARIMAYFGGK 506
DB 84 GPIMLDEVCTGTEASLADCKSLGWLKSNCRHREDAGVC 123
OY 507 GPIHVDNVACTGNERSLADCIKQDIGHNCRHSEDAGVIC 546

3
US-08-473-791-10 STANDARD; PRT: 585 AA.

Sequence 10, Application US/08473791
Patent No. 5736340
GENERAL INFORMATION:
APPLICANT: Koths, Kirston E.
APPLICANT: Halenbeck, Robert F.
APPLICANT: Taylor, Eric W.
APPLICANT: Wang, Alice M.
APPLICANT: Casipit, Clayton L.
TITLE OF INVENTION: Secreted Mac-2-Binding Glycoprotein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cetus Oncology Corporation
STREET: 1400 Fifty-Third Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,791
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,714
FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Kenneth M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 2595.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 420-3152
TELEFAX: (510) 658-5470
TELEX: N/A
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 585 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

SQ SEQUENCE 585 AA; 65330 MW; 1874129 CN;
 Query Match 51.7%; Score 391; DB 1; Length 585;
 Best Local Similarity 52.0%; Pred. No. 6,80e-30;
 Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

24 VRLADGATNGRVEIFRGQMGTVCDNLMDLTDASVCRALGFENATOLGRAAFGGGS 83
 447 VRLMDGENKKEGRVEIFNGQMGTVCDNLMDLTDASVCRALGFENATOLGRAAFGGGS 506
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

84 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 123
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

T 4
 5510466-4 STANDARD: PRT: 491 AA.

AC xxxxxx
 DT 01-JAN-1900
 XX Patent No. 5510466.
 DE Patent No. 5510466.
 CC Patent No. 5510466
 CC APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO
 CC TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY
 CC THERETO
 CC NUMBER OF SEQUENCES: 12
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/307,400
 CC FILING DATE: 16-SEP-1994
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 997,113
 CC FILING DATE: 24-DEC-1992
 CC APPLICATION NUMBER: 391,486
 CC FILING DATE: 09-AUG-1989
 CC APPLICATION NUMBER: 272,002
 CC FILING DATE: 15-NOV-1988
 CC SEQ ID NO:4
 CC LENGTH: 453
 CC SEQUENCE 491 AA; 54317 MW; 1350575 CN;

Query Match 49.7%; Score 376; DB 4; Length 453;
 Best Local Similarity 49.0%; Pred. No. 2.46e-28;
 Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

352 VRLVGGSPHGRVIEFHGQMGTVCDRLWELRGLVVCRLGYKQVSHKRAYFGKGT 411
 447 VRLMDGENKKEGRVEIFNGQMGTVCDNLMDLTDASVCRALGFENATOLGRAAFGGGS 506
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

412 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 451
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

RESULT 5
 ID US-08-794-795-7 STANDARD: PRT: 489 AA.

AC xxxxxx
 DT
 XX Sequence 7, Application US/08794795
 CC Patent No. 5916766
 CC GENERAL INFORMATION:
 CC APPLICANT: Elshouraghy, Nab11
 CC APPLICANT: Adamou, John
 CC APPLICANT: Gross, Mitchell
 CC APPLICANT: Lyoko, Paul
 CC TITLE OF INVENTION: Human Macro Scavenger Rec

CC TITLE OF INVENTION: eptor
 CC NUMBER OF SEQUENCES: 9
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: SmithKline Beecham Corporation
 CC STREET: 709 Swedeland Road
 CC CITY: King of Prussia
 CC STATE: PA
 CC COUNTRY: USA
 CC ZIP: 19406
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM Compatible
 CC OPERATING SYSTEM: DOS
 CC SOFTWARE: FASTSEQ for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/794,795
 CC FILING DATE: 04-FEB-1997
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 0ATG50009P
 CC FILING DATE: 22-MAY-1996
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Han, William T
 CC REGISTRATION NUMBER: 34,344
 CC REFERENCE/DOCKET NUMBER: ATG50009
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 610-270-5219
 CC TELEFAX: 610-270-4026
 CC TELEX:
 CC INFORMATION FOR SEQ ID NO: 7:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 489 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 489 AA; 49441 MW; 1141296 CN;

Query Match 48.7%; Score 369; DB 2; Length 489;
 Best Local Similarity 54.0%; Pred. No. 1.31e-27;
 Matches 54; Conservative 17; Mismatches 24; Indels 5; Gaps 2;

394 VRLNGGTNR--GRAEYVYNNMGITICDDWDNDATVFCRLGY--SRGRLSSYGGGS 448
 447 VRLMDGENKKEGRVEIFNGQMGTVCDNLMDLTDASVCRALGFENATOLGRAAFGGGS 506
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

449 GNTWLDNVNCRGTENSLMDCSKNSMGHNCVHNEDAGVEC 488
 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVYC 546

RESULT 6
 ID US-08-392-367B-2 STANDARD: PRT: 518 AA.

AC xxxxxx
 DT
 XX Sequence 2, Application US/08392367B
 CC Patent No. 5691197
 CC GENERAL INFORMATION:
 CC APPLICANT: Tryggvason, Karl
 CC APPLICANT: Elioma, Outi
 CC APPLICANT: Kangas, Maarit
 CC TITLE OF INVENTION: An isolated DNA Sequence for a
 CC TITLE OF INVENTION: No. 5691197e1 Macrophage Receptor with
 CC TITLE OF INVENTION: a Collagenous Domain and the
 CC TITLE OF INVENTION: Polypeptide Chain Encoded by
 CC TITLE OF INVENTION: such a Sequence
 CC NUMBER OF SEQUENCES: 2

```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fay, Sharpe, Beall, Fagan,
CC ADDRESSEE: Minnich & McKee
CC STREET: 1100 Superior Avenue
CC STREET: Suite 700
CC CITY: Cleveland
CC STATE: Ohio
CC COUNTRY: U.S.A.
CC ZIP: 44114-2518
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch,
CC MEDIUM TYPE: 720 Kb storable
CC COMPUTER: IBM PS/2, Model 35 SX
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: Word Perfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/392,367B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Minnich, Richard J.
CC REGISTRATION NUMBER: 24,175
CC REFERENCE/DOCKET NUMBER: TRV 2 009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (216) 861-5582
CC TELEFAX: (216) 241-1666
CC TELEX: (216) 980162
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 518 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: Single
CC TOPOLOGY: Linear
CC SEQUENCE 518 AA: 52730 MW; 1281000 CN;
SO
Dd 423 VRIMGQNR--GRAEYVNNENGTICDDDDNNDATYFCRMIGY--SRGALSSYGGS 477
Qy 447 VRLMDGENKKGRVEVFINGOMGTICDDGMDTDDKDAIVICRLGKGPARTMAYFEGK 506
Dd 478 GNIMLDNVCRCGTENSLMDCKSNNGNNCVCHNEAGIEC 517
Qy 507 GPIHDVNVKCTGENERSLADCIKODIGRHNCRHSEDAVIC 546
RESULT 7 STANDARD; PRT; 451 AA.
ID US-08-154-365-2
xxxxx
xx DE Sequence 2, Application US/08154365
xx CC Sequence 2, Application US/08154365
xx CC Patent No. 5624904
xx CC GENERAL INFORMATION:
xx CC APPLICANT: Duane, Dana W.
xx CC APPLICANT: Resnick, David
xx CC APPLICANT: Kreiger, Monty
xx CC APPLICANT: Joiner, Keith A.
xx CC TITLE OF INVENTION: Method for Treating Gram-Positive
xx CC TITLE OF INVENTION: Septicemia
xx CC NUMBER OF SEQUENCES: 2
xx CC CORRESPONDENCE ADDRESS:
xx CC ADDRESSEE: Patrea L. Pabst
xx CC STREET: 1100 Peachtree Street, Suite 2800
xx CC City: Atlanta
xx CC STATE: Ga
xx CC COUNTRY: USA

```

```

CC      ZIP: 30309-4530
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/154,365
CC      FILING DATE:
CC      CLASSIFICATION: 514
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Pabst, Patrea L.
CC      REGISTRATION NUMBER: 31,284
CC      REFERENCE/DOCKET NUMBER: MIT392
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (404)-815-6558
CC      TELEFAX: (404)-815-6555
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 451 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      HYPOTHEetical: YES
CC      ANTI-SENSE: NO
CC      FRAGMENT TYPE: N-terminal
CC      ORIGINAL SOURCE:
CC      ORGANISM: homo sapien
CC      PUBLICATION INFORMATION:
CC      AUTHORS: Ashkenas, et al.
CC      JOURNAL: J. Lipid Res.
CC      VOLUME: 34
CC      PAGES: 983-1000
CC      DATE: 1993
CC      RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 451
SQ      SEQUENCE 451 AA; 49841 MW; 1025052 CN;

Query Match          48.3%; Score 366; DB 1; Length 451;
Best Local Similarity 49.0%; Pred.No. 2,68e-27;
Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

Dd   350 VRLVGGSPHGGRVEILHSGQWGTICDDNMEVRGQYVVCSSLGIPIGVQAAYKKAHFEQGT 409
    |||:| :|||||: :||| | | | : : ||| | | | : : ||| |
Qy   447 VRLDNGENKKRGRAVEFLINGQWGTICDDGWTDKDAAVICRQLGYGPARRATYAIFEGEK 506

Db   410 GPIWLNEVFCEGFRESSIEECKIRONGTRACSHSEDAGVTC 449
    |||::| | | | | | | | | | | | | | | | | | | | | |
Qy   507 GPIHDNVKCTGNERSLADCIKQDIGRHNCRHSESDAGVTC 546

RESULT      8
xx   US-08-453-117-2        STANDARD:          PRI:       451 AA.
xx      xxxxxx
xx      DT
xx      DE
xx      Sequence 2, Application US/08453117
cc      Patent No. 5683903
cc      GENERAL INFORMATION:
cc      APPLICANT: Lyeko, Paul G.
cc      APPLICANT: Elshourbagy, Nahil A.
cc      APPLICANT: Brawner, Mary E.
cc      TITLE OF INVENTION: Attachment Enhanced 293 Cells
cc      NUMBER OF SEQUENCES: 4
cc      CORRESPONDENCE ADDRESS:
cc      ADDRESSEE: SmithKline Beecham - Corporate Patents
cc      ADDRESSEE: U.S.
cc      STREET: Mailcode - UW2220, 709 Swedeland Road
cc      CITY: King of Prussia
```


CC STATE: Pennsylvania
 CC COUNTRY: U.S.A.
 CC ZIP: 19406-5090
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.30
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/453,117
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Jervais, Herbert H.
 CC REGISTRATION NUMBER: 31,171
 CC REFERENCE/DOCKET NUMBER: SBC-P50338
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (610) 270-5019
 CC TELEFAX: (610) 270-5090
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 451 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;
 SQ
 Query Match 47.7% Score 361; DB 1; Length 451;
 Best Local Similarity 49.0%; Pred. No. 8.84e-27;
 Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;
 DB 350 VRLVGGSGPHEGRVEILHSGQWGTICDDREVEVRVGVCRSLGYPGQAVHKAHFGOGT 409
 |||:| |||||:||||||| | : |||| | | : |||||
 QY 447 VRLMDGNNKKEGRVEYFINGQWGTICDDGWTDKDAVAICQLGKGPARRATMAFGECK 506
 |||:| |||||:||||||| | : |||| | | : |||||
 DB 410 GPIWLNEVFCFGRESSIECKIRQWGTTRACSHSDEAGVTC 449
 |||:| |||||:||||||| | : |||| | | : |||||
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
 |||:| |||||:||||||| | : |||| | | : |||||
 RESULT 9
 ID US-08-973-145-2 STANDARD: PRT: 451 AA.
 XX
 AC xxxxxx
 XX
 Sequence 2, Application US/08973145
 CC Sequence 2, Application US/08973145
 CC Patent No. 5919636
 CC GENERAL INFORMATION:
 CC APPLICANT: Lysko, Paul G.
 CC APPLICANT: Elshourbagy, Nabil A.
 CC APPLICANT: Bramer, Mary E.
 CC TITLE OF INVENTION: Attachment Enhanced 293 Cells
 CC NUMBER OF SEQUENCES: 4
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US96/08081
 CC FILING DATE:
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Jervais, Herbert H.
 CC REGISTRATION NUMBER: 31,171
 CC REFERENCE/DOCKET NUMBER: P50338
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (610) 270-5019
 CC TELEFAX: (610) 270-5090
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 451 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;
 SQ

CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Elizabeth J. Hecht
 CC REGISTRATION NUMBER: P-41, 824
 CC REFERENCE/DOCKET NUMBER: P50338
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (610) 270-5009
 CC TELEFAX: (610) 270-5090
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 451 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;
 SQ
 Query Match 47.7% Score 361; DB 2; Length 451;
 Best Local Similarity 49.0%; Pred. No. 8.84e-27;
 Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;
 DB 350 VRLVGGSGPHEGRVEILHSGQWGTICDDREVEVRVGVCRSLGYPGQAVHKAHFGOGT 409
 |||:| |||||:||||||| | : |||| | | : |||||
 QY 447 VRLMDGNNKKEGRVEYFINGQWGTICDDGWTDKDAVAICQLGKGPARRATMAFGECK 506
 |||:| |||||:||||||| | : |||| | | : |||||
 DB 410 GPIWLNEVFCFGRESSIECKIRQWGTTRACSHSDEAGVTC 449
 |||:| |||||:||||||| | : |||| | | : |||||
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
 |||:| |||||:||||||| | : |||| | | : |||||
 RESULT 10
 ID PCT-US96-08081-2 STANDARD: PRT: 451 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 Sequence 2, Application PC/TUS9608081
 DE
 XX
 CC Sequence 2, Application PC/TUS9608081
 CC GENERAL INFORMATION:
 CC APPLICANT: SmithKline Beecham Corporation
 CC TITLE OF INVENTION: Attachment Enhanced 293 Cells
 CC NUMBER OF SEQUENCES: 4
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: SmithKline Beecham - Corporate Patents U.S.
 CC STREET: Mailcode - UW2220, 709 Swedeland Road
 CC CITY: King of Prussia
 CC STATE: Pennsylvania
 CC COUNTRY: U.S.A.
 CC ZIP: 19406-5090
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.30
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US96/08081
 CC FILING DATE:
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Jervais, Herbert H.
 CC REGISTRATION NUMBER: 31,171
 CC REFERENCE/DOCKET NUMBER: P50338
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (610) 270-5019
 CC TELEFAX: (610) 270-5090
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 451 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 451 AA; 49762 MW; 1024670 CN;
 SQ

DE Sequence 2, Application US/08794795
XX
CC Sequence 2, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC APPLICANT: Eishourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 495 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 495 AA: 49764 MW: 1218428 CN:
Query Match 43.1%; Score 326; DB 2; Length 495;
Best Local Similarity 48.5%; Pred. No. 3,61e-23;
Matches 49; Conservative 18; Mismatches 27; Indels 7; Gaps 5;
DB 399 VRIV-GSSNR-GRAEYVYSTGTICDDQNSDAIVFCRMIGY---SKRAL-YKVGAG 452
QY 447 VRLMDENKKEGRVEYFINGQWGTICDDGWTDXDAVVICRQLGKGPARTWAY-FGEG 505
DB 453 TGOIWLNDVQCRGTESTIMSCITKNSMGHHDCSHEEDAGVGC 493
QY 506 KGPIDHNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
RESULT 14
ID US-08-794-795-6 STANDARD: PRT: 520 AA.
XX
AC xxxxxx
XX
DE Sequence 6, Application US/08794795
XX
CC Sequence 6, Application US/08794795
CC Patent No. 5916766
CC GENERAL INFORMATION:
CC

CC APPLICANT: Eishourlagy, Nabil
CC APPLICANT: Adamou, John
CC APPLICANT: Gross, Mitchell
CC APPLICANT: Lyoko, Paul
CC TITLE OF INVENTION: Human Macro Scavenger Rec
CC TITLE OF INVENTION: eptor
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FASTSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/794,795
CC FILING DATE: 04-FEB-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: ATG50009P
CC FILING DATE: 22-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Han, William T
CC REGISTRATION NUMBER: 34,344
CC REFERENCE/DOCKET NUMBER: ATG50009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-5219
CC TELEFAX: 610-270-4026
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 520 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 520 AA: 52658 MW: 1340662 CN:
Query Match 43.1%; Score 326; DB 2; Length 520;
Best Local Similarity 48.5%; Pred. No. 3,61e-23;
Matches 49; Conservative 18; Mismatches 27; Indels 7; Gaps 5;
DB 424 VRIV-GSSNR-GRAEYVYSTGTICDDQNSDAIVFCRMIGY---SKRAL-YKVGAG 477
QY 447 VRLMDENKKEGRVEYFINGQWGTICDDGWTDXDAVVICRQLGKGPARTWAY-FGEG 505
DB 478 TGOIWLNDVQCRGTESTIMSCITKNSMGHHDCSHEEDAGVGC 518
QY 506 KGPIDHNVKCTGNERSLADCIKODIGRHNCRHSEDAVVIC 546
RESULT 15
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.
XX
AC xxxxxx
XX
DE Sequence 2, Application US/08200900A
XX
CC Sequence 2, Application US/08200900A
CC Patent No. 5665566
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
CC

CC STREET: 87 Cambridgepark Drive
 CC CITY: Cambridge
 CC STATE: MA
 CC COUNTRY: USA
 CC ZIP: 02140
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/200,900A
 CC FILING DATE: 23-FEB-1994
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Meinerdt, Maureen C.
 CC REGISTRATION NUMBER: 31,544
 CC REFERENCE/DOCKET NUMBER: GI 5201-FWC
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 876-1170 X8574
 CC TELEFAX: (617) 876-5851
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 798 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 798 AA: 88937 MW: 3292434 CN;
 SQ
 Query Match 16.58; Score 125; DB 1; Length 798;
 Best Local Similarity 30.68; Pred. No. 2.50e-03;
 Matches 19; Conservative 13; Mismatches 28; Indels 2; Gaps 2;
 Db 457 VRLFNGTSSGIQVPRISIMHVACAEWNTQISDDVQCLGL-GNGNS-SVPTFTGG 514
 Qy 447 VRLMDCENKREGRVEYFINGQWGTICDDGWTDAVAICRQLGKGPARRIMAYFEEGK 506
 Db 515 GP 516
 Qy 507 GP 508

Search completed: Mon Mar 13 10:35:28 2000
 Job time : 7 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
 Run on: Mon Mar 13 10:34:56 2000; MasPar time 6.76 Seconds
 Tabular output not generated. 353.990 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (447-547) from US09147947A.pep (6 of 6)
 Perfect Score: 757
 Sequence: 1 VRLMDGENKKEGRVEVFING.....KODIGRHNCRHSEDAGVICD 101

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseqs35
 1:geneseqp

Statistics: Mean 28.103; Variance 102.859; scale 0.273

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	822	1 W99087	Human serine protease	3.21e-73
2	757	100.0	875	1 W83361	Human neutrotrypsin	3.21e-73
3	724	95.6	761	1 W83362	Mouse neutrotrypsin	1.97e-69
4	724	95.6	761	1 W99088	Mouse serine protease	1.97e-69
5	391	51.7	585	1 W81363	Human cytokine regulat	1.08e-31
6	391	51.7	585	1 R36533	GP85-97 clone 18 prod.	1.08e-31
7	391	51.7	585	1 R40172	Sequence of an immunor	1.08e-31
8	391	51.7	585	1 R41339	Tumour associated 90K	1.08e-31
9	391	51.7	585	1 R40213	Sequence of a 90K tumo	1.08e-31
10	376	49.7	453	1 R05509	Scavenger receptor pro	5.02e-30
11	376	49.7	453	1 R27036	Bovine sol. scavenger	5.02e-30
12	369	48.7	489	1 W39749	Mouse macroSR protein	3.00e-29
13	369	48.7	518	1 W03561	Macrophage receptor w1	3.00e-29
14	366	48.3	451	1 W19708	Macrophage scavenger r	6.45e-29
15	366	48.3	451	1 W64591	Human SRP protein.	6.45e-29
16	361	47.7	451	1 R27035	Human sol. scavenger r	2.31e-28
17	361	47.7	451	1 W08077	Type I macrophage scav	2.31e-28
18	361	47.7	451	1 R40802	Human scavenger recept	2.98e-28
19	360	46.9	451	1 R21512	Human scavenger recept	1.07e-27
20	355	46.2	666	1 W64590	Human SRP protein fra	3.80e-27
21	350	45.7	1290	1 W07609	Rat von Ebner's gland	1.05e-26
22	346	45.7	574	1 R40784	Sequence of cyclophilin	1.05e-26
23	346	45.7	574	1 W81362	Mouse cytokine regulat	1.05e-26

ALIGNMENTS

24	332	43.9	347	1 Y13369	Amino acid sequence of	3.67e-25
25	332	43.9	347	1 W68200	Human scavenger recept	3.67e-25
26	332	43.9	347	1 W64537	Human liver cell clone	3.67e-25
27	326	43.1	495	1 W39747	Human macroSR protein	1.68e-24
28	326	43.1	520	1 W39748	Human macroSR protein	1.68e-24
29	326	43.1	520	1 Y04374	Human macrophage recep	1.68e-24
30	326	43.1	520	1 Y04374	Human macrophage recep	1.68e-24
31	312	16.5	798	1 W57283	Bovine enterokinase.	2.25e-03
32	312	16.5	798	1 W57283	Bovine enterokinase.	2.25e-03
33	82	11.2	359	1 W35849	Human CD5 for use in T	1.14e+00
34	82	11.2	359	1 W35849	Human CD5 for use in T	1.14e+00
35	82	11.2	359	1 W35849	Human CD5 for use in T	1.14e+00
36	82	11.2	359	1 W35849	Human CD5 for use in T	1.14e+00
37	82	11.2	359	1 W35849	Human CD5 for use in T	1.14e+00
38	79	10.4	1167	1 W87632	Bacillus thuringiensis	3.45e+01
39	79	10.4	1167	1 W87632	Bacillus thuringiensis	3.45e+01
40	77	10.2	1162	1 W17700	Cyert4.	3.45e+01
41	77	10.2	1162	1 W06829	Cyert4.	3.45e+01
42	76	10.0	1162	1 P70137	Major neutralizing ant	5.07e+01
43	76	10.0	1162	1 P70137	Major neutralizing ant	5.07e+01
44	76	10.0	1162	1 R59841	Sequence of a region o	6.14e+01
45	76	10.0	1162	1 R59841	Sequence of a region o	6.14e+01
					Human APOE4L.	6.14e+01
					Human APOE4Lx2.	6.14e+01
					ApoE4Lx2 protease.	6.14e+01

At No 6 is available but from

RESULT 1

ID	W99087	standard; Protein; 822 AA.
AC	W99087	
DI	13-MAY-1999	(first entry)
DE	Human serine protease BSSP-3.	
KW	Serine protease; BSSP-3; brain tissue.	
OS	Homo sapiens.	
PN	W09905280-A1.	
PD	04-FEB-1999.	
PF	24-JUL-1998; J03324.	
PR	24-JUL-1997; JP-213969.	
PA	(SUNR) SUNTORY LTD.	
PI	Tsuruoka N, Yamaguchi N, Yamashiro K;	
DR	WPI: 99-142942/12.	
DR	N-PSDB; X19024.	
PT	New serine protease expressed in brain tissue - used in screening	
PS	for potential serine protease inhibitors for drug use	
PS	Claim 1: Page 61-65; 69pp; Japanese.	
CC	The present sequence is a serine protease designated BSSP-3, which	
CC	is isolated from human brain tissue. Transformants may be used to	
CC	produce the enzyme or its partial sequences. Products from the present	
CC	invention are used for screening for potential peptide or non-peptide	
CC	serine protease inhibitors or expression regulators for use as drugs.	
SQ	Sequence 822 AA:	

Query Match 100.0%; Score 757; DB 1; Length 822;
 Best Local Similarity 100.0%; Pred. No. 3.21e-73;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	447 VRLMDGENKKEGRVEVFINGOMGTTICDDGMDKDAVIGCRGKPARPMAYGECK 506
QY	447 VRLMDGENKKEGRVEVFINGOMGTTICDDGMDKDAVIGCRGKPARPMAYGECK 506
DB	507 GPIHVDNVCCTGNERSLADICIKODIGRHNCRHSEDAGVICD 547
QY	507 GPIHVDNVCCTGNERSLADICIKODIGRHNCRHSEDAGVICD 547

RESULT 2

ID	W83361	standard; Protein; 875 AA.
AC	W83361	
DI	17-FEB-1999	(first entry)
DE	Human neutrotrypsin.	
KW	Human; neutrotrypsin; tumour inhibition; neurological disease;	
KW	lung disease; gene therapy; drug development; stroke; brain injury;	
KW	neurodegeneration; neuroinflammatory disease; multiple sclerosis;	
KW	epilepsy; hypoxia; ischaemia; nerve transection; neuroangiogenesis;	

KW emphysema; bronchitis.
 OS Homo sapiens.
 PN WO9849322-A1.
 PD 05-NOV-1998.
 PR 24-APR-1998: IB0625.
 PF 26-APR-1997: CH-000966.
 PI (SOND/) SONDEREGGER P.
 DR Sandergerger P;
 WI: 99-009438/01.
 N-PSDB: V72589.
 PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 20-24; 50pp: English.
 CC The present sequence represents human neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases; e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,
 CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 875 AA:

Query Match 100.0%; Score 757; DB 1; Length 875;
 Best Local Similarity 100.0%; Pred. No. 3.21e-73;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 500 VRLMDGNKKEGVEVFVINGQMTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 559
 QY 447 VRLMDGNKKEGVEVFVINGQMTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 560 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 600
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 547

RESULT 3
 ID W81363 standard; Protein: 761 AA.
 AC W81363.
 DT 17-FEB-1999 (first entry)

DE Mouse neurotysin.
 PT Mouse; neurotysin; tumour inhibition; neurological disease;
 PT lung disease; gene therapy; drug development; stroke; brain injury;
 PT neurodegeneration; neuroinflammatory disease; multiple sclerosis;
 PT epilepsy; hypoxia; ischaemia; nerve transection; neovascularisation;
 PT emphysema; bronchitis.

OS Mus musculus.
 PN WO9849322-A1.
 PD 05-NOV-1998.
 PR 24-APR-1998: IB0625.
 PF 26-APR-1997: CH-000966.
 PI (SOND/) SONDEREGGER P.
 DR Sandergerger P;
 WI: 99-009438/01.
 N-PSDB: V72590.

PT New human and murine neurotysin - used, e.g. for inhibiting
 PT tumours, treatment of neurological or lung disease, including by
 PT gene therapy and in drug development
 PS Claim 1: Page 29-32; 50pp: English.
 CC The present sequence represents mouse neurotysin. Neurotysin proteins
 CC and polynucleotides can be used: (i) to inhibit tumours, including
 CC metastases; e.g. of brain or retina; (ii) to minimise tissue damage
 CC caused by stroke or brain injury (having a protective effect on the
 CC penumbra zone); (iii) to treat or prevent neurodegeneration,

CC neuroinflammatory disease (e.g. multiple sclerosis) or epilepsy; (iv) to
 CC increase survival of damaged neurons (e.g. in cases of hypoxia,
 CC ischaemia, nerve transection) and to stimulate regeneration and/or
 CC restoration of synapses; (v) to treat or prevent retinal disorders (e.g.
 CC degeneration or neovascularisation); (vi) to prevent apoptosis (or other
 CC causes of cell death) in the nervous system; (vii) to regenerate brain
 CC and/or nervous tissue; (viii) to treat pain; (ix) to improve brain
 CC performance, including learning and memory; (x) to treat or prevent a
 CC wide range of psychiatric disorders; and (xi) to treat brain or lung
 CC injury associated with protease expression (specifically emphysema or
 CC bronchitis).
 SQ Sequence 761 AA:

Query Match 95.6%; Score 724; DB 1; Length 761;
 Best Local Similarity 92.1%; Pred. No. 1.97e-69;
 Matches 93; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 386 IRLVDGNKKEGVEVFVINGQMTICDDGWTDKHAAVICRQLGYGPARARTMAFGECK 445
 QY 447 VRLMDGNKKEGVEVFVINGQMTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 486
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 547

RESULT 4
 ID W99088 standard; Protein: 761 AA.
 AC W99088.
 DT 13-MAY-1999 (first entry)

DE Mouse serine protease BSSP-6.
 KW Serine protease; BSSP-3; brain tissue.
 OS Mus sp.
 PN WO9905290-A1.
 PD 04-FEB-1999.
 PR 24-JUL-1998: J03324.
 PF 24-JUL-1997: JP-213969.
 PA (SUNR) SUNTOY LTD.
 PI Tsuruoka N, Yamaguchi N, Yamashiro K;
 DR WPI: 99-142942/12.
 N-PSDB: X19027.

PT New serine protease expressed in brain tissue - used in screening
 PT for potential serine protease inhibitors for drug use
 PS Example 1: Page 51-54; 69pp: Japanese.
 CC The present sequence is a serine protease designated BSSP-3, which
 CC is isolated from mouse brain tissue. Transforms may be used to
 CC produce the enzyme or its partial sequences. Products from the present
 CC invention are used for screening for potential peptide or non-peptide
 CC serine protease inhibitors or expression regulators for use as drugs.
 SQ Sequence 761 AA:

Query Match 95.6%; Score 724; DB 1; Length 761;
 Best Local Similarity 92.1%; Pred. No. 1.97e-69;
 Matches 93; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

DB 386 IRLVDGNKKEGVEVFVINGQMTICDDGWTDKHAAVICRQLGYGPARARTMAFGECK 445
 QY 447 VRLMDGNKKEGVEVFVINGQMTICDDGWTDKDAVICRQLGYGPARARTMAFGECK 506
 DB 446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 486
 QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSSEDAVYICD 547

RESULT 5
 ID W81363 standard; Protein: 585 AA.

AC W81363;
 DT 12-APR-1999 (first entry)
 DE Human cytokine regulatory factor CYTRF.
 KW Cytokine regulatory factor; cyclophilin C-associated protein;
 KW CYTRF; CYCAP; antiinflammatory; inflammation; autoimmune disease;
 KW multiple sclerosis; rheumatoid arthritis; ankylosing spondylitis;
 KW meningitis; herpes encephalitis; viral meningoencephalitis;

KW	viral hepatitis; graft rejection; graft versus host disease;
OS	allergy; asthma; diagnosis; therapy; human.
KW	Homo sapiens.
FT	Key
FT	Location/Qualifiers
FT	Misc_difference 257
FT	/note= "encoded by CGr"
PN	W09856819-A1.
PD	17-DEC-1998.
PF	12-JUN-1998; U12345.
PR	13-JUN-1997; US-049578.
PA	(STRD) UNIV LELAND STANFORD JUNIOR.
PI	Conboy IB, Jones PP;
DR	WPI: 99-080890/07.
DR	NPDB: V68491.
DR	Reduction of pro-inflammatory cytokines in a T cell population - using cytokine regulatory factor (CYRF), useful in the diagnosis and treatment of pro-inflammatory conditions
DR	·Disclosure: Page 47-51; 77pp; English.
CC	This is the amino acid sequence of human cytokine regulatory factor (CYRF), also known as cyclophilin C-associated protein (CYCAP).
CC	Methods are provided for the modulation of cytokine production by T helper cells using CYRF (especially human CYRF).
CC	composition comprising CYRF as an active agent is administered in vitro or in vivo, and can act on mature, committed Th1 type T cells to decrease the production of pro-inflammatory cytokines, or to skew the commitment of precursor T helper cells to Th1 or a Th1 cytokine profile. The administration of CYRF is useful in the diagnosis and treatment of autoimmune diseases (e.g. multiple sclerosis, rheumatoid arthritis, and degenerative joint diseases including ankylosing spondylitis and reactive arthritis), and inflammation caused by bacterial, viral or parasitic infection (e.g. meningitis, herpes encephalitis, viral meningoencephalitis and viral hepatitis), including response to vaccination (e.g. vaccines containing rabies, varicella zoster, and measles) and skin sensitivity (e.g. graft rejection and graft vs host disease). CYRF signalling and related pathways are also useful for modelling and screening new pharmacological agents. The administration of agents that block CYRF action (e.g. antibodies) are useful in reducing unwanted allergic responses, especially in asthma.
CC	Sequence 585 AA;

```

Query Match          51.7%; Score 391; DB 1; Length 585;
Best Local Similarity 52.0%; Pred. No. 1,08e-31;
Matches          52; Conservative          17; Mismatches          31; Indels          0; Gaps          0;

24 MRLADGGATNOGRREIIFYRGOWGVCDNLMDLTDASVSCRALGPENATOALGRAAGGGS 83
   || || :|||:|: |||||:|:|: ||:|:|:|:|: ||:|:|:|:|: ||:|:|:|:|:
447 VRLMDGKKEKREGRVEITINGOWGITICDGTDKDAAYICQIQLGKGPAPARTMAVFEBGK 506

Db
84 GPIMLDEVOCITGTEASLADCKSLGMLKSNCRHERHDAGVC 123
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 507 GPIHVDNWKCTGNGERSLADCTIKDIGHNCRHSDAAGVIC 546

RESULT
6
ID R36533 standard; Protein; 585 AA.
AC R36533;
DT 19-AUG-1993 (first entry)
DF GP85-97 Clone 18 prod.
DR Glycoprotein; lectin; Mac-2; PHA; lymphocytes; leucoagglutinating;
KW Glycohemagglutinin; cancer; amplification; PCR.
OS Homo sapiens.
FH Key
FT peptide location/Qualifiers
   1..18
   /note="leader sequence"
   19..585
   /note="mature gp85-97"

PN WO9308215-A.
PD 29-APR-1993.
PE 15-OCT-1992; U08878.
PR 16-OCT-1991; US-777121.
PR 15-OCT-1992; US-961404.
PA (CETU ) CETUS ONCOLOGY CORP.

```

PI Casipit CL, Halenbeck R, Kotis KE, Taylor EM, Wang AM:
DR WPI: 93-152424/18.
PT New glyco:protein complex binding to human lectin Mac-2 - also
PT interferes with PHA activation of Lymphocytes for treating and
PT preventing cancer, infectious diseases, etc.
PS Claim 1, Page 41; 57pp; English.
CC SK-BR-3 gp97 recovered in partially proteolyzed form was denatured
CC and reduced and the 97 and 70 kd mols. were purified using size
CC exclusion HPLC in 0.1 percent SDS. The 97 and 70 kd mols. were
CC digested with Lys-C protease and the resulting peptides purified and
CC sequenced. The N-terminal sequence of gp97 was used to design
CC degenerate primers for use in a PCR reaction on SK-BR-3 mRNA. PCR
CC reactions using primer 2 in combination with primers 1 and 3 gave prods.
CC of 97 and 121 bp. Further PCR using primers 2, 3 and 4 yielded DNA
CC sequences of 740 and 765 respectively. A DNA sequence was obtd. from
CC the PCR prods. Two additional oligonucleotide sequences were
CC synthesised based on this sequence and used to probe a Tth-1 CDNA
CC library to obtain the full-length cDNA sequence that encodes SK-BR-3
CC gp97. Both strands of clone 218 were sequenced. It encodes a novel
CC glycoprotein binds the Mac-2 lectin and interferes with PHA activation
CC of lymphocytes. It can be used for treating or preventing diseases that
CC result from binding of a disease-causing agent to the cell surface of a
CC target cell. The gp. can be used in treatment of cancers, partic.
CC breast cancer. See also R36532.
SQ Sequence 585 AA;

Query Match	51.7%	Score 391;	DB 1;	Length 585;
Best Local Similarity	52.0%;	Pred. No. 1,08e-31;		
Matches 52;	Conservative 17;	Mismatches 31;	Indels 0;	Gaps 0;

Db	24	MRADGCGATNGRGREIIFRYGCGMGVCCNLDLIDASVYCAALGFENATOLGPAAGGGS	83
	:		:
QY	447	VRLDGKGKKEGREGVEYFINGGWFICDDGTBDXDAAYICRQLOLYKGPARRATMAVYGEKG	506
	:		
Db	84	GPIMLDEVQCTGTASLADCKSLGMLKSNRHRERDAVYC	123
	:		
QY	507	GPIMVNVKTCIGNERSLADCIKODIGRHNCRHSEDADVIC	546
	:		

RESULT	7
ID	R40172 standard; Protein: 585 AA.
AC	R40172;
DT	08-FEB-1994 (first entry)
DE	Sequence of an immunoregulin (IR)-95 polypeptide.
KW	Tumour associated antigen; breast cancer; cell line CG-5;
OS	ovarian cancer; immunoregulin-95; IR-95.
OS	Homo sapiens.
EH	Key
FT	peptide
FT	Location/Qualifiers
FT	1..17
FT	/label= signal
FT	24..125
FT	/label= SRCR homology region
FT	modified_site
FT	71
FT	/label= glycosylation site
FT	/note= "see also AAs 125,192,362,398,551,580"
PN	W09316180-A.
PD	19-AUG-1993.
PF	17-FEB-1993; E00382.
PR	17-FEB-1992; IT-RM0100.
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA	(UYCH) UNIV CHERIF ANNINO G D.
PI	Azam M, Iacobelli S, Natoli C, Sures I, Ullrich A;
PI	WPL: 93-272884/34.
DR	N-PSDB; Q46888.
DR	Recombinant DNA - encoding tumour associated antigen,
PT	Immunoregulin-95
PT	Immunoregulin-95
PS	Disclosure; Fig 1; 69pp; English.
CC	Immunoregulin-95 is a 90k tumour-associated antigen purified from
CC	the culture fluid of the human breast cancer cell line, CG-5, the
CC	serum of a breast cancer patient, or the ascitic fluid from an
CC	ovarian cancer patient. The native antigen, which has a mol. wt. of
CC	95 kD, is present as a high mol. wt. complex. Homology in the region
CC	of AAs 35-80 of the 90k antigen is found with type I macrophage

CC scavenger receptor: sea urchin speract receptor: and human
CC lymphocyte glycoprotein TI/Len-1.
SQ Sequence 585 AA;

Query Match 51.7%: Score 391; DB 1; Length 585;
Best Local Similarity 52.0%: Pred. No. 1.08e-31;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGQMGTVCDNLMDLTDAASYCRALGFENATQALGRAAFGQGS 83
Y 447 VRLMDGKNGKRGREVEIFNGQMGTCDDGWTDKAAVTCRQLGYKGPARRATMAFGECK 506

DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123
Y 507 GPIHDVNVKCTGNERSLADCKIKODIGRHNCRHSEDAGVC 546

RESULT 8
R41359 standard: Protein; 585 AA.
R41359; 03-MAR-1994 (first entry)
Tumour associated 90K antigen.
Antigen: cancer: inflammation; autoimmune disease; viral infection.
Homo sapiens.

OS Key Location/Qualifiers
FT peptide 1. .18
FT /label= signal peptide.
FT region 24. .125
FT /note= "Region homologous with sea urchin speract receptor"

FT modified_site 69
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 125
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 192
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 362
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 398
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 551
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site 580
FT /note= "Potential asparagine linked N-glycosylation site"

FT modified_site

W09317119-A.

02-SEP-1993.

15-FEB-1993; E00379.

17-FEB-1992; IT-RM0099.

PA (UYCH-) UNIV CHIETI ANNUNZIO G D.

PA (UYNY) UNIV NEW YORK STATE.

PI Iacobelli S, Natoli C, Schlessinger J;

DR WPI: 93-288423/36.

DR N-PSDB: Q46161.

PT 90K tumour-associated antigen - purified from serum of breast

PT cancer patient or ascitic fluid from ovarian cancer patient

PS Claim 1; Figure 1; 73pp; English.

CC The purified 90K antigen or its antigenic determinant containing

CC fragment can be assayed to diagnose a disorder, such as cancer (

CC especially breast or ovarian cancer) or a viral infection,

CC inflammation, autoimmune disease and/or arthritis in a patient.

CC It may also be used in the preparation of an agent for the

CC treatment of these disorders.

CC Sequence 585 AA;

Query Match 51.7%: Score 391; DB 1; Length 585;
Best Local Similarity 52.0%: Pred. No. 1.08e-31;

Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGQMGTVCDNLMDLTDAASYCRALGFENATQALGRAAFGQGS 83

Y 447 VRLMDGKNGKRGREVEIFNGQMGTCDDGWTDKAAVTCRQLGYKGPARRATMAFGECK 506

DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123

Y 507 GPIHDVNVKCTGNERSLADCKIKODIGRHNCRHSEDAGVC 546

RESULT 9
R40213 standard: Protein; 585 AA.
R40213; 03-FEB-1994 (first entry)
Sequence of a 90K tumour-associated antigen, IR-95, capable
DE of binding to SP-2 (Accession No. I-1083).
KM Tumour-associated antigen; breast cancer cell line CG-5;
KM 90 K antigen.
OS Homo sapiens.

OS Key Location/Qualifiers
FT peptide 1. .18
FT /label= signal
FT region 24. .125
FT /label= SCRC homology region
FT binding_site 69
FT /label= potential glycosylation site
FT /note= "see also AAs 125,192,363,398,551,580"

PN W09316181-A.

PD 19-AUG-1993; E00385.

PF 17-FEB-1993; E00385.

PA 17-FEB-1992; IT-RM0099.

PA (UYCH-) UNIV CHIETI ANNUNZIO G D.

PA (UYNY) UNIV NEW YORK STATE.

PI Iacobelli S, Natoli C, Schlessinger J;

DR WPI: 93-272885/34.

DR N-PSDB: Q48170.

PT New 90K tumour-associated antigen, IR-95 - is for use in

PT diagnosis and therapy of cancer, HIV and auto-immune diseases

PS Claim 2; pages 45-48; 68pp; English.

CC 90K antigen has an apparent molecular weight of approx. 95 kd. It is

CC a tumour associated antigen and is elevated in the serum of patients

CC with cancer and also in patients with HIV. It reacts with Mab SP-2

CC which was prod. by immunising mice with proteins that had been

CC released into tissue culture fluid by human MCF-7 breast cancer

CC cells. Mab SP-2 cell line is deposited at the Institut Pasteur,

CC Paris. Accession number 1-1083. 90K is also present in normal

CC subjects. It is purified from the culture fluid of the human breast

CC cancer cell line; CG-5. N-terminal sequencing of the 90K antigen

CC was used to design a 'guesser' nucleotide sequence (Q48171) as a

CC probe to screen a lambda-gt10 library pred. from MCF7 polyA+ RNA.

CC The complete nucleotide sequence of isolated clones is given in

CC Q48170.

CC Sequence 585 AA;

Query Match 51.7%: Score 391; DB 1; Length 585;
Best Local Similarity 52.0%: Pred. No. 1.08e-31;
Matches 52; Conservative 17; Mismatches 31; Indels 0; Gaps 0;

DB 24 MRLADGATNGRVEIFRYGQMGTVCDNLMDLTDAASYCRALGFENATQALGRAAFGQGS 83

Y 447 VRLMDGKNGKRGREVEIFNGQMGTCDDGWTDKAAVTCRQLGYKGPARRATMAFGECK 506

DB 84 GPIMDEVQCTGTETASLADCKSLGWLKSNCRHERDAGVC 123

Y 507 GPIHDVNVKCTGNERSLADCKIKODIGRHNCRHSEDAGVC 546

RESULT 10
R05509 standard: protein; 453 AA.
R05509; 23-OCT-1990 (first entry)
Scavenger receptor protein with affinity for acylated low density

DE

DE lipoprotein (aLDL). lipoprotein: aLDL; atherosclerotic plaque;
KW Acylated low density lipoprotein: aLDL; atherosclerotic plaque;
KM ds.
PN WO9005748-A.
PD 31-MAY-1990.
PF 14-NOV-1989; 005116.
PR 15-NOV-1988; US-272002.
PA (MAST) Massachusetts Inst Tech.
PI Engellenner TJ;
DR WPI: 90-193408/25.
DR N-PSDB: 004926.
PT New receptor protein -
PT has affinity for acetylated low density lipoprotein and
PT corresponding antibodies and DNA sequences.
PT Disclosure: 79pp; English.
CC Receptor protein, and fragments and analogues thereof may be
CC immobilised on a support and used in assay and purification of the
CC aLDL target. Labelled Abs, raised in assay and purification of the
CC into the vascular system to detect the presence of atherosclerotic
CC plaques.
SQ Sequence 453 AA:

Query Match 49.7%; Score 376; DB 1; Length 453;
Best Local Similarity 49.0%; Pred. No. 5.02e-30;
Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

DB 352 VRLVGGSGPHEGRVFEHFGOMGTVCDDRMELKGLVCRSLGKGVSHKRAYFGKGT 411
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
QY 447 VRLMDGNKKRGVEVFINGOMGTICDDGWTDKDAVITCROLGKGPARTAYFGEKG 506
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
DB 412 GPIWLNEVFCFPRESSIEECRIKQWGVACSHEDAGVTC 451
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVIC 546
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||

RESULT 11
ID R27036 standard; Protein: 453 AA.
AC R27036;
DT 16-FEB-1993 (first entry)
DE Bovine sol. scavenger receptor.
KW Macrophage; endotoxaemia; radiolabelled; toxic; degenerate.
OS Bos taurus.
FH Key
FT domain Location/Qualifiers
FT 1..50
FT /note="cytoplasmic"
FT 51..76
FT /note="transmembrane"
FT 77..109
FT domain
FT /note="spacer"
FT 110..271
FT domain
FT /note="alpha helical coiled coil"
FT 272..343
FT /note="collagen binding"
FT 341..451
FT /note="Cys-rich"
FT domain
PN WO9214482-A.
PD 03-SEP-1992.
PF 21-FEB-1992; U01370.
PR 22-FEB-1991; US-662227.
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
PI Krieger M;
DR WPI: 92-315935/38.
DR N-PSDB: 028540.
PT Treatment and diagnosis of endotoxaemia and related disease
PT states using a polypeptide fragment of the extracellular
PT portion of a macrophage scavenger receptor protein
PS Disclosure: Page 38; 56pp; English.
CC The bovine scavenger receptor protein sequence was deduced from the
CC DNA sequence obt. by screening a bovine lung cDNA library with
CC degenerate probes designed based on fragments of the purified
CC scavenger receptor protein. The scavenger receptor protein has a
CC binding capacity for acetylated low density lipoprotein (LDL) and
CC binds endotoxin. Fragments of the extracellular portion of the

CC scavenger receptor protein inactivate endotoxin-related substances
CC and are used in the diagnosis and treatment of endotoxaemia. They
CC may be administered to patients at high risk of symptomatic or
CC endotoxic shock. The fragments may also be fixed to inert supports
CC for purification purposes. The fragments may be modified to have
CC greater binding affinity for the endotoxin-related substance than
CC the native scavenger receptor protein, or to more effectively
CC neutralise the toxic or pathogenic effects of mols. that bind
CC the scavenger receptor protein or of organisms which express such mols.
CC See also R27035.
SQ Sequence 453 AA:

Query Match 49.7%; Score 376; DB 1; Length 453;
Best Local Similarity 49.0%; Pred. No. 5.02e-30;
Matches 49; Conservative 17; Mismatches 34; Indels 0; Gaps 0;

DB 352 VRLVGGSGPHEGRVFEHFGOMGTVCDDRMELKGLVCRSLGKGVSHKRAYFGKGT 411
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
QY 447 VRLMDGNKKRGVEVFINGOMGTICDDGWTDKDAVITCROLGKGPARTAYFGEKG 506
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
DB 412 GPIWLNEVFCFPRESSIEECRIKQWGVACSHEDAGVTC 451
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
QY 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVIC 546
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||

RESULT 12
ID W39749 standard; Protein: 489 AA.
AC W39749;
DT 20-MAY-1998 (first entry)
DE Mouse macrophage protein.
KW Macrophage scavenger receptor; macrophage; mouse; antibodies; treatment;
KW diagnosis; cardiovascular disease; macrophage; bacterial infection;
KW immune cell related host cell disorder; pancreatitis; organ failure;
KW endotoxaemia; agonist; antagonist; dysfunction.
OS Mus sp.
PN EP-808899-A2.
PD 26-NOV-1997.
PF 19-MAY-1997; 303394.
PR 04-FEB-1997; US-794795.
PR 23-MAY-1996; US-017699.
PA (SMK) SMITHKLINE BEECHAM CORP.
PI Adamou JE, Elshourbagy NAE, Gross MS, Lysko PG;
DR WPI: 98-001786/01.
PT Human macrophage scavenger receptor - used to develop products for
PT treating e.g. cardiovascular disorders, septic shock, bacterial
PT infections etc.
PS Disclosure: Fig 1; 45pp; English.
CC This sequence represents a mouse macrophage scavenger receptor (MMacroSR)
CC which is used to confirm the presence of a novel human macrophage
CC scavenger receptor (HMacroSR). This polypeptide and resulting antibodies can be
CC used to treat cardiovascular diseases, including atherosclerosis,
CC hypertension myocardial and cerebral infarction, angina, organ failure,
CC stroke, gangrene, and loss of function in the extremities. They can also
CC be used to treat or diagnose various macrophage and other immune cell
CC related host defense disorders, septic shock, pancreatitis, multiple
CC organ failure, endotoxaemia and infections caused by gram negative and
CC gram positive bacteria. Agonists and antagonists may also be used to
CC treat these diseases. The polynucleotide can be used to detect mutated
CC forms of HMacroSR which may be associated with a dysfunction and can
CC define a diagnosis of a disease (or susceptibility to a disease)
CC resulting from altered HMacroSR expression.
SQ Sequence 489 AA:

Query Match 48.7%; Score 369; DB 1; Length 489;
Best Local Similarity 54.0%; Pred. No. 3.00e-29;
Matches 54; Conservative 17; Mismatches 24; Indels 5; Gaps 2;

DB 394 VRLMGSTNR--GRAEYVNNMGITICDDMDNDVAFVCRMGY--SGRALSYGGGS 448
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
QY 447 VRLMDGNKKRGVEVFINGOMGTICDDGWTDKDAVITCROLGKGPARTAYFGEKG 506
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||
DB 449 GNIWLDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVIC 488
|||:| |||||:|||||:|||||:|:|:| ||||| | |||||

QY 507 GPIHVDNVKCTGNERSLADCIKQDIGRHNCRHSEDAGVIC 546

Sequence	Accession	Protein	Function
1	W03561	standard; Protein; 518 AA.	
2	W03561		
3	22-JAN-1997	(first entry)	
4	Macrophage receptor with collagenous domain (MARCO).		
5	Macrophage receptor-collagenous; MARCO; immune defense;		
6	phagocytosis.		
7	Mus sp.		
8	Key		
9	domain		
10	FT	Location/Qualifiers	
11	FT	1. 49	
12	FT	/label= Domain-I	
13	FT	/note= "hydrophilic domain not containing a signal	
14	FT	peptide"	
15	FT	50. 74	
16	FT	/label= Domain-II	
17	FT	/note= "hydrophobic domain"	
18	FT	75. 149	
19	FT	/label= Domain-III	
20	FT	/note= "hydrophilic domain"	
21	FT	87. 89	
22	FT	/label= Glycosylation	
23	FT	/note= "putative N-glycosylation site"	
24	FT	138. 140	
25	FT	/label= Glycosylation	
26	FT	/note= "putative glycosylation site"	
27	FT	150. 419	
28	FT	/label= Domain-IV	
29	FT	/note= "collagenous structure characterised by 89	
30	FT	Gly-X-X repeats with one interruption	
31	FT	(Ala-Glu-Lys at 174-176)"	
32	FT	420. 518	
33	FT	/label= Domain-V	
34	FT	/note= "globular C-terminal domain"	
35	FT		
36	PD	W09626219-A1.	
37	PD	29-AUG-1996.	
38	PD	19-FEB-1996; F10091	
39	PR	21-FEB-1995; US-392367.	
40	PA	(ELON//) ELONMAA O.	
41	PA	(KANG//) KANGAS M.	
42	PA	(TRYG//) TRYGGVASON K.	
43	PI	Elomaa O, Kangas M, Tryggvason K;	
44	PI	WPI: 96-402332/40.	
45	DR	N-PSDB: T39340.	
46	PT	New macrophage receptor, MARCO, that binds bacteria but not yeast	
47	PT	and related DNA, anti-sense probes and antibodies, involved in	
48	PT	immune defence and phagocytosis	
49	PS	Claim 10; Fig 2A-B; 62PP; English.	
50	CC	A novel macrophage receptor with a collagenous domain, named MARCO	
51	CC	(W03561), is a membrane-bound trimeric protein which is expressed	
52	CC	strongly after birth in a subset of macrophages in mouse spleen and	
53	CC	lymph nodes. It binds to bacteria but not to yeast, and probably	
54	CC	plays a role in immune defence and/or phagocytosis. The amino acid	
55	CC	sequence of MARCO was deduced from DNA (see also T39340) obtd. from	
56	CC	a mouse macrophage cDNA library. Recombinant MARCO can be produced	
57	CC	in transformed host cells. It is used to raise antibodies useful	
58	CC	in detecting MARCO or interfering with its function.	
59	CC	Sequence 518 AA;	

	Query Match	48.7%	Score 369;	DB 1;	Length 518;
	Best Local Similarity	54.0%;	Pred. No.	3,00e-29;	
Matches	54; Conservative		17; Mismatches	24; Indels	5; Gaps 2;
D b	423 VRINGGTR--GRAEVYNNEMGGTCDDDMNNAATVCRMLGY---SRGRALSSYGGS		477		
O y	447 VRLDNGENKKEGRVEFTNIGQWGTICDDGWTDKAAYACROLRGAPARARTMAFYEGEK		506		
D b	478 GNIMLDNVNCRTGENSLMDCSKNSMGVCNCHNEADAGEC		517		
O y	507 GPIHDVNKYCTGNERSLADCIKKODIGRNCHNSEDAGYIC		546		

RESULT 14
ID W19708 standard; Protein; 451 AA.
AC W19708;
DT 19-AUG-1997 (first entry)
DE Macrophage scavenger receptor protein.
KW Macrophage scavenger receptor protein; MSRP; human; lipoteichoic acid;
KW LTA; Gram-positive bacteria; cell wall; atherosclerosis; host defence;
KW septicæmia; inhibitor; complement activation; cytokine release; therapy;
KW nitric oxide production; bacterial infection; septic shock.
OS Homo sapiens.
PN US5624904-A.
PD 29-APR-1997.
PT 17-NOV-1993;
PR 17-NOV-1993; US-154365.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PA (UYYA) UNIV YALE.
PI Joiner KA, Krieger M;
DR WPI: 97-258336/23.
DR N-PSDB: T68796.
PT Treatment of septicæmia caused by Gram-positive bacteria - by
PT administration of macrophage scavenger receptor protein
PS Disclosure: Column 23-26: 16pp: English.
CC This sequence represents the human macrophage scavenger receptor protein
CC (MSRP). MSRP binds specifically to the lipoteichoic acid (LTA) residues
CC on the Gram-positive bacterial cell wall. MSRP exhibit unusually broad
CC binding specificity for polyanionic ligands, and have been implicated in
CC atherosclerosis and a variety of host defence functions. This protein can
CC be used in the method of the invention. The method of the invention is
CC for the treatment of septicæmia caused by Gram-positive bacteria. The
CC method comprises administration of a MSRP in an amount sufficient to
CC inhibit complement activation, cytokine release or nitric oxide
CC production induced by LTA released by the Gram-positive bacteria. In
CC addition to treating Gram-positive septic shock, the MSRP can be used to
CC screen for other compounds for treating Gram-positive septic shock. MSRP
CC can also be used to purify, label or detect LTA or LTA-containing cells.
CC MSRP, active MSRP fragments, anti-MSRP antibodies or other compounds that
CC inhibit binding of Gram-positive bacteria to MSRP can be used to treat
CC pathologies such as septicæmia, Gram-positive bacterial infection,
CC Gram-positive septicæmia or Gram-positive or Gram-negative septic shock.
SQ Sequence 451 AA;

	Query Match	48.3%	Score 366	DB 1	length 451
	Best Local Similarity	49.0%	Pred. No. 6.4e-29		
	Matches	49	Conservative	17	Mismatches 34; Indels 0; Gaps 0
Db	350	VRLVGGSGPHGRVETILSGOMGTICDDNVEVRGVYCRSLGYPGVQAVKAAHFGGT	409		
			:::		
Oy	447	VRLMDGKKKRGREVEFLNGOMGTICDDGWTDKAAAYICROLGKGPARRATMAYFEQK	506		
Db	410	GPIMLNEVFCFGRESSIEECKIRPMGTACRACSHSDACVTC	449		
		:::			
Oy	507	GPIMVDNKCCTGNERSLADCLIKODIGRHNCRHSDAGAYIC	546		

RESULT 15
ID W64591 standard; Protein; 1785 AA.
AC W64591;
DT 23-OCT-1998 (first entry)
KW Human SRCR protein.
KW Scavenger receptor, cysteine rich domain, SRCR, diagnosis; treatment;
KW nervous system, medullo-blastoma; glioma; breast; detection;
OS autoantibody; ss.
OS Homo sapiens.
PN W09830687-A2.
PD 16-JUL-1998.
PE 09-JAN-1998; D00096.
PR 18-JUL-1997; DE-030997.
PR 09-JAN-1997; DE-000519.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI Mollenhauer J, Poustka A;
DR WPI; 98-399136/34.
DR N-PSDB; V49652.

THIS PAGE BLANK (USPTO)

SCRC4
947-6-06.rpr
Use #4 No. 3

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

search_pp protein - protein database search, using Smith-Waterman algorithm
Run on:      Mon Mar 13 10:34:27 2000;  MasPar time 10.39 Seconds
Tabular output not generated.  458.750 Million cell updates/sec

```

```

Title: >US-09-147-947-6
Description: (447-547) from US09147947A.pep (6 of 6)
Perfect Score: 757
Sequence: 1 VRLDGKNEKKEGRNEVFING.....KODIGRHNCRHSDEAGVTCDD 101

```

Scoring table: PAM 150

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:      pir62
               1:pir1 2:pir2 3:pir3 4:pir4
Statistics:    Mean 37.980;  Variance 62.138;  scale 0.6111
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Accession No.	Score	Query Length	DB	ID	Description	Pred. No.
1	724	95.6	761.1	2	JC5759	2
2	466	61.6	213.3	2	T14893	2
3	391	51.7	35.5	2	A41761	2
4	385	50.9	43.3	2	S08276	1
5	381	50.3	44.4	2	I46862	2
6	380	50.2	111.6	2	S36077	2
7	380	50.2	114.9	2	I38006	2
8	380	50.2	115.1	2	I38004	2
9	380	50.2	113.6	2	I38005	2
10	373	49.3	44.4	2	A44407	2
11	369	48.7	51.8	2	A55840	2
12	368	48.6	44.8	2	B44407	2
13	362	47.8	145.6	2	A46496	2
14	361	47.7	45.1	2	A38415	2
15	357	47.2	57.7	2	A53202	2
16	353	46.6	91.8	2	JC4361	2
17	351	46.4	33.2	2	A32751	2
18	350	46.2	129.0	2	A57190	2
19	345	45.6	60.0	2	S56744	2
20	343	45.3	50.4	2	S56745	2
21	323	42.7	46.8	2	S26741	2
22	320	42.3	66.6	2	T45100	2
23	157	20.7	103.4	1	A53563	2
					brain-specific serine scavenger receptor cy	2
					Mac-2-binding glycopr	2
					LDL receptor I, macro	2
					macrophage scavenger	2
					M130 antigen - human	2
					M130 antigen (extrac	2
					M130 antigen (cytosol	2
					M130 antigen (cytosol	2
					macrophage scavenger	2
					macrophage bacteric	2
					antigen WCL1 precurs	2
					macrophage scavenger	2
					cyclophilin C-associ	2
					scavenger receptor Cy	2
					spleen receptor prec	2
					eberrin precursor - r	2
					mucin (clone pgm7-1)	2
					mucin (clone pgm31-1)	2
					T-cell glycoprotein C	2
					muscD precursor - mou	2
					enterepoelipase (PC 3	2

	RESULT	6	S36077	#type complete	
	ENTRY TITLE		M130 antigen - human		
	ORGANISM		#formal_name Homo sapiens #common_name man		
	DATE		06-Jan-1995 #sequence_revision 06-Jan-1995	#text_change	
			02-Aug-1996		
	ACCESSIONS REFERENCE		I38003; S36077		
	#authors		Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.		
	#journal		Eur. J. Immunol. (1993) 23:2320-2325		
	#title		A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.		
	#cross-references	WJID:93380506			
	#accession	I38003			
	##status		Preliminary; translated from GB/EMBL/DBJ		
	##molecule-type	mRNA			
	##residues	1-1116	#label RES		
	##cross-references	EMBL:Z29686; NID:g3j2141; PID:g3j2142			
	CLASSIFICATION FEATURE		#superfamily scavenger receptor cysteine-rich domain homology		
		43-147	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC1\		
		151-254	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC2\		
		258-361	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC3\		
		365-468	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC4\		
		470-573	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC5\		
		575-678	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC6\		
		711-814	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC7\		
		816-920	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC8\		
		921-1024	#domain scavenger receptor cysteine-rich domain homology		
			#label SRC9		
	SUMMARY		#length 1116	#molecular-weight 120979	#checksum 5687
	Query Match		50.2%;	Score 380;	DB 2; Length 1116;
	Best Local Similarity	51.0%;	Pred. No. 3,48e-65;		
	Matches	51; Conservative	19; Mismatches 30;	Indels 0;	Gaps 0;
Db	714	LRIVGGRCAGVEYLYHSGSWTICDDSDMDSDAIVVCROLGCEATNATGSAHFGEGT	773		
Oy	447	VRLMDGNKKEGVPEFINGOMTCITDDBDTDKDAVICRQIGRYGPAPARIMAYFGESK	506		
Db	774	GPWLDEMCNGKESRIWOCHSHGWQNCRRHKEDAGVIC	813		
Oy	507	GPILHDVVKCTGNERSLADCIKODIRHNCRHSESDAGVIC	546		
RESULT ENTRY	7				
TITLE		I38006	#type complete		
ORGANISM		M130 antigen (#intracellular variant) - human			
DATE		17-May-1996 #sequence_revision 17-May-1996	#text_change		
		02-Aug-1996			
ACCESSIONS REFERENCE		I38006; S35768			
#authors		I38003			
#journal		Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.			
#title		Eur. J. Immunol. (1993) 23:2320-2325			
#cross-references	WJID:93380506				
#accession	I38006				
##status		Preliminary; translated from GB/EMBL/DBJ			
##molecule-type	mRNA				

```

##residues 1-1149 ##1label RES
##cross-references EMBL:Z229571; NID:9312147; PID:9312148
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
FEATURE
43-147 #domain scavenger receptor cysteine-rich domain homology
151-254 #label SRC1\ #domain scavenger receptor cysteine-rich domain homology
258-351 #label SRC2\ #domain scavenger receptor cysteine-rich domain homology
355-468 #label SRC3\ #domain scavenger receptor cysteine-rich domain homology
470-573 #label SRC4\ #domain scavenger receptor cysteine-rich domain homology
608-711 #label SRC5\ #domain scavenger receptor cysteine-rich domain homology
744-847 #label SRC6\ #domain scavenger receptor cysteine-rich domain homology
849-953 #label SRC7\ #domain scavenger receptor cysteine-rich domain homology
9-1057 #label SRC8\ #domain scavenger receptor cysteine-rich domain homology
MAY #label SRC9 #domain scavenger receptor cysteine-rich domain homology
#length 1149 #molecular-weight 124328 #checksum 487

Query Match 50.2% Score 380; DB 2; Length 1149;
Best Local Similarity 51.0% Pred. No.3,48e-65;
Matches 51; Conservative 19; Mismatches 30; Indels 0; Gaps 0;

Db 747 LRLVNGGRCAGAEVEIYHEGSMGTICDDSDMLSDAHVVCROLGCGEAINMGASHFEGT 806
:||||: ||||: |||||: || |||||: || |||||: || |||||: || |||||: || |||||:
Oy 447 VRLMGENKKEGVEVFINGQWGTICDDGTDTDAATVCOLGKGRARATMAVFEGR 506

Db 807 GPIWDEMKKCNKESRIWQCHSHGQAOQCRHKEDAVIC 846
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Oy 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVIC 546

RESULT 8
ENTRY #type complete
TITLE I38004
ORGANISM MI30 antigen (cytosolic variant 1) - human
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
07-Feb-1997
ACCESSIONS I38004; S36078
REFERENCE I38003
AUTHORS Law, S.K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.;
Willits, A.C.; Mason, D.Y.;
Eur. J. Immunol. 1993 23:2320-2325
A new macrophage differentiation antigen which is a member of
the scavenger receptor superfamily.
#journal #title
#cross-references MUID:93380506
#status I38004
#molecule_type mRNA
#residues 1-1151 ##1label RES
#cross-references EMBL:Z229561; NID:9312143; PID:9312144
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS cytosol
FEATURE
43-147 #domain scavenger receptor cysteine-rich domain homology
151-254 #label SRC1\ #domain scavenger receptor cysteine-rich domain homology
258-351 #label SRC2\ #domain scavenger receptor cysteine-rich domain homology
355-468 #label SRC3\ #domain scavenger receptor cysteine-rich domain homology
470-573 #label SRC4\ #domain scavenger receptor cysteine-rich domain homology
575-678 #label SRC5\ #domain scavenger receptor cysteine-rich domain homology
711-814 #label SRC6\ #domain scavenger receptor cysteine-rich domain homology
#label SRC7\

```

Query Match	Best Local	Similarity	51.0%	Score	380.	DB 2:	Length	1151:	Mismatches	19:	Mismatches	30:	Indels	0:	Gaps	0:
816-920	#domain scavenger receptor cysteine-rich domain homology	#label SRC8\	#domain scavenger receptor cysteine-rich domain homology	#label SRC9	#length 1151	#molecular-weight 124820	#checksum 1481									
921-1024																
SUMMARY																
Db	714	LRVYGGRCAGRYEITHESGWTICDDSDWSDSDAHVVCROLCGEAINAGSAHFEGET	773													
QY	447	VRMDGKKRGREYEVINGOMTICDDGDTWKDAVIRQLGTYGPARARTMAVFEGR	506													
Db	774	GPVWDEKCKGKESRIMQCHSHGQOQCRKEDAGVIC	813													
QY	507	GPVHDVVKCTGNERSLADCIKODIRHNCRHSDEAGVIC	546													
RESULT	9															
ENTRY		I38005	#type complete													
TITLE		M130 antigen (cytosolic variant 2) - human														
ORGANISM		#formal name Homo sapiens #common name man														
DATE		17-May-1996 #sequence_revision 17-May-1996 #text_change 07-Feb-1997														
ACCESSIONS		I38005; S36079														
REFERENCE		I38003														
#authors		Law, S.-K.; Micklem, K.J.; Shaw, J.M.; Zhang, X.P.; Dong, Y.; Willis, A.C.; Mason, D.Y.														
#journal		Eur. J. Immunol. (1993) 23:2320-2325														
#title		A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily.														
#cross-references		MOTID:93380506														
#accession		I38005														
#status		preliminary; translated from GB/EMBL/DBJ														
#molecule_type		mRNA														
#residues		1-1156	#label RES													
#cross-references		EMBL:Z22970; NID:g312145; PID:g312146														
CLASSIFICATION		#superfamily scavenger receptor cysteine-rich domain homology														
KEYWORDS		cytosol														
FEATURE																
43-147		#domain scavenger receptor cysteine-rich domain homology														
151-254		#domain SRC1\														
258-361		#domain scavenger receptor cysteine-rich domain homology														
365-468		#domain SRC2\														
470-573		#domain scavenger receptor cysteine-rich domain homology														
575-678		#domain SRC3\														
711-814		#domain scavenger receptor cysteine-rich domain homology														
816-920		#domain SRC4\														
921-1024		#domain scavenger receptor cysteine-rich domain homology														
SUMMARY		#domain SRC5\														
		#domain scavenger receptor cysteine-rich domain homology														
		#domain SRC6\														
		#domain scavenger receptor cysteine-rich domain homology														
		#domain SRC7\														
		#domain scavenger receptor cysteine-rich domain homology														
		#domain SRC8\														

RESULT 10
ENTRY A44407 #type complete
TITLE macrophage scavenger receptor, RSRL (collagen-like domain) -
ORGANISM rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
#label rabbit
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
16-Jul-1999
ACCESSIONS A44407
#authors DOI, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.;
Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.;
Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
J. Biol. Chem. (1993) 268:2126-2133
#journal
#file Charged collagen structure mediates the recognition of
negatively charged macromolecules by macrophage scavenger
receptors.
#cross-references M01D:93131972
#accession A44407
#status preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-454 ##label DOI
#experimental_source lung
#note sequence extracted from NCBI backbone (NCBIP:123206)
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
cysteine-rich domain homology
FEATURE 350-453
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 454 #molecular-weight 49735 #checksum 2435
Query Match 49.3%; Score 373; DB 2; Length 454;
Best Local Similarity 48.0%; Pred. No. 1,67e-63;
Matches 48; Conservative 20; Mismatches 32; Indels 0; Gaps 0;
DB 353 VRLVGRGPHGREGVEILANGOMGLVCDHMLRAGQVYCRSLGYKRVKRAVFGQGT 412
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 447 VRLMDGNNKRGREYEVINGQWGTICDDGWTDKDAAYICRLQGLGKPARATMAYFEGR 506.
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
DB 413 GPTWLANEPCLGMSISIECKIRQWGVYVCSHGSDAGVTC 452
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 507 GPIHVDNVKCTGNERSLADICIKDIGHNCRHSDAGVTC 546
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
ULT 11
#type complete
#formal_name Mus musculus #common_name house mouse
#label SRC
DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change
16-Jul-1999
ACCESSIONS A55840
#authors Elomaa, O.; Kangas, M.; Sahlberg, C.; Tuukkanen, J.;
A55840
#journal Tryggvason, K.
#title Cell (1995) 80:603-609
#file Cloning of a novel bacteria-binding receptor structurally
related to scavenger receptors and expressed in a subset of
macrophages.
#cross-references M01D:95171455
#accession A55840
#status preliminary
#molecule_type mRNA
#residues 1-518 ##label ELO
#cross-references GB:U8424; NID:9682722; PID:9682723
CLASSIFICATION #superfamily scavenger receptor cysteine-rich domain homology
KEYWORDS transmembrane protein
FEATURE 418-518
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 518 #molecular-weight 52730 #checksum 8159

Query Match 48.7%; Score 369; DB 2; Length 518;
Best Local Similarity 54.0%; Pred. No. 1.51e-62;
Matches 54; Conservative 17; Mismatches 24; Indels 5; Gaps 2;
DB 423 VRINGGTRN--GRAEYVNNEMGTCDDMDNDATYFCMLGT---SRGALSTYGGGS 477
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 447 VRLMDGNNKRGREYEVINGQWGTICDDGWTDKDAAYICRLQGLGKPARATMAYFEGR 506
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
DB 478 GNTMDNVNCRGTENSLMDCSKNSMGWNCVHNEDAGVEEC 517
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
OY 507 GPIHVDNVKCTGNERSLADICIKDIGHNCRHSDAGVTC 546
|||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:| |||||:|
RESULT 12
ENTRY B44407 #type complete
TITLE macrophage scavenger receptor, RSRL - mouse
ALTERNATE_NAMES macrophage scavenger receptor type I
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
20-Sep-1999
ACCESSIONS B44407; A38260; I56334
#authors DOI, T.; Higashino, K.; Kurihara, Y.; Wada, Y.; Miyazaki, T.;
Nakamura, H.; Uesugi, S.; Imanishi, T.; Kawabe, Y.;
Itakura, H.; Yazaki, Y.; Matsumoto, A.; Kodama, T.
J. Biol. Chem. (1993) 268:2126-2133
#journal
#file Charged collagen structure mediates the recognition of
negatively charged macromolecules by macrophage scavenger
receptors.
#cross-references M01D:93131972
#accession B44407
#status preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues 1-458 ##label DOI
#experimental_source macrophage-like cell line P38BD
#note sequence extracted from NCBI backbone (NCBIP:123207)
REFERENCE A38260
#authors Freeman, M.; Ashkenas, J.; Rees, D.J.G.; Kingsley, D.M.;
Copeland, N.G.; Jenkins, N.A.; Krieger, M.
Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8810-8814
#journal An ancient, highly conserved family of cysteine-rich protein
#title domains revealed by cloning type I and type II murine
macrophage scavenger receptors.
#cross-references M01D:91062370
#accession A38260
#status preliminary
#molecule_type mRNA
#residues 349-458 ##label FRE
#cross-references GB:M59445; GB:W36817; NID:g192737; PIDN:AAA37464.1;
PID:g192738
REFERENCE I56334
#authors Ashkenas, J.; Penman, M.; Vasile, E.; Acton, S.; Freeman,
M.W.; Krieger, M.
J. Lipid Res. (1993) 34:983-1000
#journal Structures and high and low affinity ligand binding
#title properties of murine type I and type II macrophage
scavenger receptors.
#cross-references M01D:93359822
#accession I56334
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 5-458 ##label RES
#cross-references GB:L04274; NID:g2933745; PIDN:AAA39747.1; PID:g2933746
CLASSIFICATION #superfamily unassigned collagens; scavenger receptor
cysteine-rich domain homology
FEATURE 354-457
#domain scavenger receptor cysteine-rich domain homology
#label SRC
SUMMARY #length 458 #molecular-weight 50130 #checksum 1435
Query Match 48.6%; Score 368; DB 2; Length 458;
Best Local Similarity 49.0%; Pred. No. 2.63e-62;
Matches 49; Conservative 15; Mismatches 36; Indels 0; Gaps 0;

```

Db      357 VRLVGGSGAHEVEIFHOQOMGICDDRDMDIRAGOVCRSLGYOEVLAVKKRHHFGGT    416
|||::| | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      447 VRLMDEENKKEGVYEFINQMKTICDGDGMTDKRAANVICRDLGKGPARATMAFYGECK    506
||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db      417 GPIWLEVMCFRGRESSIENCKINMGVLSCSHSDAGVTC    456
||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      507 GPIHDVNVCCTGNEISRADCIKODIGRHNCRHSSEDAVYC    546
||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT   13
ENTRY     A46496 #type complete
TITLE     antigen WC1.1 precursor - bovine
ORGANISM  #formal_name Bos primigenius taurus #common_name cattle
DATE      18-Jun-1993 #sequence_revision 19-May-1994 #text_change
          13-Nov-1998
ACCESSION A46496; I45834; S19913
REFERENCE A46496
#authors  WJjngard, P.L.; Metzelaar, M.J.; Machugh, N.D.; Morrison,
           W.I.; Clevers, H.C.
journal   J. Immunol. (1992) 149:3273-3277
title     Molecular characterization of the WC1 antigen expressed
           specifically on bovine CD4-CD8- gamma delta T lymphocytes.
#cross-references MIMD:93056489
#accession A46496
#status preliminary
#molecule_type mRNA
#residues 1-1436 ##label WIJ
##cross-references EMBL:X63723; NID:g13; PID:g14
##experimental_source CD4-CD8- gamma delta T lymphocytes
#note sequence extracted from NCBI backbone (NCBIP:117475)
I45834
#authors  WJjngard, P.L.; Machugh, N.D.; Metzelaar, M.J.; Romberg, S.;
           Bensaïd, A.; Pepin, L.; Davis, W.C.; Clevers, H.C.
#journal  J. Immunol. (1994) 152:3476-3482
#title    Members of the novel WC1 gene family are differentially
           expressed on subsets of bovine CD4-CD8- gamma delta T
           lymphocytes.
#cross-references MIMD:94194107
#accession I45834
#status preliminary; translated from GB/EMBL/DDBU
##molecule_type mRNA
#residues 1-1436 ##label WI2
##cross-references EMBL:X63723; NID:g13; PID:g14
GENETICS WC1.1
#gene superfamily scavenger receptor cysteine-rich domain homology
CLASSIFICATION
FEATURE   #domain signal sequence #status predicted #label SIC\
          1-23
          24-1436 #product WC1 antigen #status predicted #label MA\
          5-131 #domain scavenger receptor cysteine-rich domain homology
          2-224 #domain scavenger receptor cysteine-rich domain homology
          236-340 #domain scavenger receptor cysteine-rich domain homology
          373-476 #domain scavenger receptor cysteine-rich domain homology
          478-581 #domain scavenger receptor cysteine-rich domain homology
          583-686 #domain scavenger receptor cysteine-rich domain homology
          687-789 #domain scavenger receptor cysteine-rich domain homology
          791-895 #domain scavenger receptor cysteine-rich domain homology
          928-1031 #domain scavenger receptor cysteine-rich domain homology
          1033-1136 #domain scavenger receptor cysteine-rich domain homology
          1152-1255 #domain scavenger receptor cysteine-rich domain homology
          #label SR11
          #length 1436 #molecular-weight 154196 #checksum 7163
SUMMARY

```

Query Match	Similarity	Score	DB 2:	Length	1436:
Best Local	53.08%	Pred. No. 7.15e-61;			
Matches	53;	Conservative	14;	Mismatches	33;
				Indels	0;
				Gaps	0;
Db	376	LRVYGGSPGCGAREVILIDDOGSWGTICDDGWDLDARVRCROLCGCEALNATGSAHFAGS	435		
Qy	447	VRLMDGENKKEGAREVEFLINGOMGTCIDGDTKDAVAIVICROLGYGPARAFRTMAVFGSG	506		
Db	436	GPIMLDNINCTGKESHWRCPSSGSGWGHCRKHODAGVTC	475		
Qy	507	GPIMVDNKCCTGNERSLADICIKODIGHNCRHSEDAGVTC	546		
RESULT	14				
ENTRY	A38415	#type complete			
TITLE	macrophage scavenger receptor splice form I - human				
ORGANISM	#formal_name Homo sapiens #common_name				
DATE	28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change				
ACCESSION	A38415				
REFERENCE	A38415				
#authors	Matsumoto, A.; Naito, M.; Itakura, H.; Ikemoto, S.; Asaoka, H.; Hayakawa, I.; Kanamori, H.; Aburatani, H.; Takaku, F.; Suzuki, H.; Kobari, Y.; Miyai, T.; Takahashi, K.; Cohen, E.H.; Wydro, R.; Housman, D.E.; Kodama, T.				
#journal	Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9133-9137				
#title	Human macrophage scavenger receptors: primary structure, expression, and localization in atherosclerotic lesions.				
#cross-references	MOL:91067661				
#accession	A38415				
#status	preliminary				
#molecule_type	mRNA				
#residues	1-451 ##label MAT				
#cross-references	GB:D90187; NID:g219989; PIDN:BAJ4208.1;				
#cross-references	PID:01014913; PID:g219990				
REFERENCE	A44408				
#authors	Emi, M.; Asaoka, H.; Matsumoto, A.; Itakura, H.; Kurihara, Y.; Wada, Y.; Kanamori, H.; Yazaki, Y.; Takahashi, E.; Lepetit, M.; Lalouel, J.; Kodama, T.; Mukai, T.				
#journal	J. Biol. Chem. (1993) 268:2120-2125				
#title	Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene.				
#cross-references	MOL:93131971				
#accession	A44408				
#status	preliminary: not compared with conceptual translation				
#molecule_type	nucleic acid				
#residues	155-872 ##label EMI				
#note	sequence extracted from NCBI backbone (NCBIP:123189)				
GENETICS					
#gene	GDB:MSR1				
##cross-references	GDB:128046; OMIM:153622				
CLASSIFICATION	#map_position 8p22-8p22				
KEYWORDS	#superfamily unassigned collagens; scavenger receptor				
FEATURE	alternative splicing; coiled coil; transmembrane protein				
347-450	#domain scavenger receptor cysteine-rich domain homology				
SUMMARY	#label SRC				
	#length 451 #molecular-weight 49762 #checksum 6168				
Query Match	47.8%;	Score 362;	DB 2:	Length 1436;	
Best Local	Similarity 49.08%;	Pred. No. 7.15e-61;			
Matches	49;	Conservative	17;	Mismatches	34;
				Indels	0;
				Gaps	0;
Db	350	VRVYGGSPGPHGAREVILHSGQWGTICDDDEWEVRCVGVCRSLGYPGVAVKHAHFGGT	409		
Qy	447	VRLMDGENKKEGAREVEFLINGOMGTCIDGDTKDAVAIVICROLGYGPARAFRTMAVFGSG	506		
Db	410	GPIMLDNINCTGKESHWRCPSSGSGWGHCRKHODAGVTC	449		
Qy	507	GPIMVDNKCCTGNERSLADICIKODIGHNCRHSEDAGVTC	546		

THIS PAGE BLANK (USPTO)

DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DE SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN PRECURSOR.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 CC Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA PANCER Z., RAST J.P., DAVIDSON E.H.;
 RT "Origins of Immunity: Transcription Factors and Effector Genes of the
 RT Vertebrate Immune System Expressed in Sea Urchin Coelomocytes.";
 RL Immunogenetics 0:0-0(1999).
 DR EMBL; AF076513; AAD05493.1; -.
 DR HSSP; P56682; ICCV.
 KW Signal; Receptor.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 1036 SCAVENGER RECEPTOR CYSTEINE-RICH PROTEIN.
 FT SEQUENCE 1036 AA; 10970704 MW; CB8BC351 CRC32;
 Query Match 54.3%; Score 411; DB 5; Length 1036;
 Local Similarity 54.0%; Pred. NO. 6,99e-76;
 Matches 54; Conservative 18; Mismatches 28; Indels 0; Gaps 0;
 Db 620 IRLVGSNEABGRVEIQTNGYWGVTICDDSGCITDASVYCRMALFGQAGSAPGSAHFGCGT 679
 QY 447 VRLMDGKRRKRGREVFETINGOMGTICDDGWDKDAVYICRLGKGPARTATMAVEFGCK 506
 Db 680 GPIOLDVGTGAGETIIFDCAPHAPEVYNCANHAIEDAGVFC 719
 QY 507 GPIHYDNYKCTGTGENSLADCTIKODIGRNCRHSSEDAGYIC 546
 RESULT 3 PRELIMINARY; PRT; 2043 AA.
 ID 096943
 AC 096943:
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
 DE SRCR DOMAIN, MEMBRANE FORM 2.
 DE SRCRM2.
 GN SRCRM2.
 OS Geodia cydonium (Sponge).
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
 CC Astrophorida; Geodiidae; Geodia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 98369060.
 RA BLUMBACH B., PANCER Z., DIEHL-SEIFERT B., STEFFEN R., MUENNER J.,
 RA MUELLER I., MUELLER W.E.G.;
 RT "The putative sponge aggregation receptor. Isolation and
 RT characterization of a molecule composed of scavenger receptor
 RT cysteine-rich domains and short consensus repeats.";
 J. Cell Sci. 111:2635-2644(1998).
 DR EMBL; Y14953; CAAT5175.1; -.
 DR HSSP; P10998; IYVC.
 DR PROSITE; PS00420; SPERACT RECEPTOR; 8.
 DR SEQUENCE 2043 AA; 220896 MW; 4CE19401 CRC32;

[illegible]

AC Q17064; 01-NOV-1996 (TREMBLrel_01, Created)
 DT 01-NOV-1996 (TREMBLrel_01, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel_12, Last annotation update)
 DE Sp85.
 OS *Arpacia punctulata* (Punctate sea urchin).
 CC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
 CC Echinozoidea; Echinacea; Arbacoidea; Arbacoidea; Arbacoidea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA DANGOTT L.J., RANKIN T.L.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RM [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89184581.
 RA DANGOTT L.J., JORDAN J.E., BELLET R.A., GARBERS D.L.;
 RT "Cloning of the mRNA for the protein that crosslinks to the egg
 RT peptide speract";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).
 DR EMBL; U30933; AAA7510.1; -.
 DR PFM; PF00530; SRCR; 4.
 DR PRINTS; PR00258; SEPARCPCPTR.
 SQ SEQUENCE 531 AA; 57414 MW; 0FE36E2D CRC32;

	Query Match	52.8%	Score 400	DB 5	Length 531
	Best Local Similarity	56.3%	Pred. No.	4,196-73	
	Matches	58	Conservative	13	Mismatches 28; Indels 4; Gaps 4
Db	261	IRIMDSGPHGGRVETIYDASGAMGTTICDGGWMAADANAAVCAKRGAKGKARRSGFGGDEFGY	320		
Oy	447	VRIMDSGKRRGRVETIYDASGAMGTTICDGGWTDKRAAVYCRIGLSTKGRPARATM-AY-FGE	504		
Db	321	TWGPHTISTVYTCGTGESSLADCVLRD-GWSSSCQHVEDAGAYVC	362		
Oy	505	GKGPHTVNDVAKCTGERSLADCTIKODIGR-INCRSESDAGAYIC	546		

RESULT	5	PRELIMINARY:	PT:	585 AA.
ID	008380			
AC	008380;			
DT	01-NOV-1996 (TEMBLrel. 01, Created)			
DT	01-NOV-1996 (TEMBLrel. 01, last sequence update)			
DT	01-NOV-1999 (TEMBLrel. 12, last annotation update)			
DE	MAC-2 BINDING PROTEIN PRECURSOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 93300818.			
RA	KOHS K., TAYLOR E., HALENBEC R., CASIPIT C., WANG A.;			
RT	"Cloning and characterization of a human Mac-2-binding protein, a new			
RT	member of the superfamily defined by the macrophage scavenger receptor			
RT	cysteine-rich domain.";			
RL	J Biol. Chem. 268:14245-14249(1993).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94308070.			
RA	ULIRICH A., SURETS I., D'EGIDO M., JALLAT B., POWELL T.J., HERBST R.;			
RA	DREPS A., AZAM M., ROBINSTEIN M., NATOLI C.;			
RT	"The secreted tumor-associated antigen 90K is a potent immune			
RT	stimulator.";			
RL	J Biol. Chem. 269:18401-18407(1994).			
DR	EMBL; L13210; AAA36193.1; -			
DR	EMBL; X79089; CAAS5699.1; -			
DR	PEAM; PF00530; SRCR; 1.			
DR	PRINTS; PR00258; SPERACTRCPTR.			
KW	Signal.			
FT	SIGNAL	1	18	POTENTIAL.
ET	CHAIN	19	585	MAC-2 BINDING PROTEIN.
QO	SEQUENCE	585 AA; 65330 MW; AAER9E32 CRC32;		

Query Match 51.7%; Score 391; DB 4; Length 585;

	Best Local Similarity	52.0%	Pred. No. 7,74e-71	Mismatches	52; Conservative	17; Mismatches	31; Indels	0; Gaps	0;
Db	24	MLRLDGGATNCGRAVEIFIRGQMGTVCDNLMDLPDAVSVCALGFEMATOALGRAAFSGS	83	::::: :	:: :	::::	:		
QY	447	VRLMGEKKRKRREVFELINGOMGITCDDGGWTDPAAYICROLGYKGPARTMTMYFEGEK	506	:	:	:	:		
Db	84	GPIMLDEVOCGTETASLADCKSLGLTKNSCNCRHERDAGVVC	123	:	:	:	:		
QY	507	GPIHDNYKCTGNERSLADCIKODGRINCHNSEBAGVTC	546	:	:	:	:		

RESULT	6	PRELIMINARY;	PRT: 1116 AA.
ID	007898		
	007898;		
	01-NOV-1996	(TREMBLrel. 01, Created)	
	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
	01-NOV-1999	(TREMBLrel. 12, Last annotation update)	
DE	M130 ANTIGEN PRECURSOR.		
GN	CD163.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
RC	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 93380506.		
RA	LAW S.A., MICKLETHAM K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,		
RA	MASON D.Y.;		
RT	"A new macrophage differentiation antigen which is a member of the		
RT	scavenger receptor superfamily.";		
RL	Eur. J. Immunol. 23:2320-2325(1993).		
RM			

RP SEQUENCE FROM N.A.
 RA RITTER M., BUECHLER C., LANGMANN T., SCHMITZ G.;
 RT "Genomic organization of the human CD163 gene";
 RL Submitted (NOV-1998) to the EMBL/genbank/DBSJ databases
 DR EMBL; Z229268; CABA054.1; -;
 DR EMBL; Y18388; CABA45233.1; -;
 DR EMBL; Y18389; CABA45233.1; JOINED.
 DR EMBL; Y18390; CABA45233.1; JOINED.
 DR EMBL; Y18391; CABA45233.1; JOINED.
 DR EMBL; Y18392; CABA45233.1; JOINED.
 DR EMBL; Y18393; CABA45233.1; JOINED.
 DR EMBL; Y18394; CABA45233.1; JOINED.
 DR EMBL; Y18395; CABA45233.1; JOINED.
 DR EMBL; Y18396; CABA45233.1; JOINED.
 DR EMBL; Y18397; CABA45233.1; JOINED.
 DR EMBL; Y18398; CABA45233.1; JOINED.
 DR EMBL; Y18399; CABA45233.1; JOINED.
 DR EMBL; Y18400; CABA45233.1; JOINED.
 DR EMBL; Y18401; CABA45233.1; JOINED.
 DR EMBL; Y18402; CABA45233.1; JOINED.
 DR EMBL; Y18403; CABA45233.1; JOINED.
 DR PFAM; PF00530; SRCR; 9.
 DR PRINTS; PRO0258; SPERACTRCPTR.
 KW Antigen; Signal;
 FT SIGNAL 1 40
 FT CHAIN 41 115 POTENTIAL.
 FT CDS 116 AA; 120979 MW; F193FBFA CRC32; M30 ANTIGEN.

[illegible]

RESULT	7	PRELIMINARY:	PRT:	1149 AA.
ID	Q07901			
AC	Q07901;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	M130 ANTIGEN, EXTRACELLULAR VARIANT PRECURSOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 93380506.			
RA	LAM S.A., MICKLEB K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,			
RA	MASON D.Y.;			
RT	"A new macrophage differentiation antigen which is a member of the			
RT	scavenger receptor superfamily.";			
RL	Eur. J. Immunol. 23:2320-2325(1993).			
DR	EMBL; Z22971; CA80544.1; -.			
DR	PFAM; PF00530; SRCR; 9.			
DR	PRINTS; PRO0258; SPERACRCPTR.			
KW	Antigen; Signal.			
FT	SIGNAL	1	40	POTENTIAL.
FT	CHAIN	1	1149	M130 ANTIGEN, EXTRACELLULAR VARIANT.
SO	SEQUENCE	1149 AA;	124328 MW;	4901CT08 CRC32;

	Query Match	50.2%	Score 380;	DB 4:	Length 1149;
	Best Local Similarity	51.0%;	Pred. No.	4,49e-68;	
	Matches	51;	Conservative	19;	Mismatches 30; Indels 0; Gaps 0;
Db	747	LRLVNGGRCAGREIYHEGS ¹ NTICDDSMLSAHAYVCROLGGGEALINATSAHPEGT	806		
OY	447	VRLDGENKKEGREVEFINQGWGTCIDDGWTDKAAYICROLGKGPARARTMAFEGEK	506		
Db	807	GPIWLDEMKCKNGKESRIINOCHSHGMOONRHKEDACIVC	846		
OY	507	GPIHYDNVKCTCGNERSLADCLIKDIGNHNCRHSDDACIVC	546		

ID	RESULT	8	PRELIMINARY:	PRF:	1151 AA.
AC	007899;				
AD	007899;				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)				
DE	M130 ANTIGEN, CYTOPLASMIC VARIANT 1 PRECURSOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 93380506.				
RA	LAW S.A., MICKLEB K.J., SHAW J.M., ZHANG X.P., DONG Y., WILLIS A.C.,				
RA	MASON D.Y.;				
RT	"A new macrophage differentiation antigen which is a member of the				
RT	scavenger receptor superfamily."				
RL	Eur. J. Immunol. 23:2320-2325(1993).				
DR	EMBL: Z22969; CAA80542.1; ..				
DR	PFAM: PF00530; SRCR: 9.				
DR	PRINTS: PRO0258; SPBRACRCPTR.				
KW	Antigen; Signal.				
FT	CHAIN	1	40	POTENTIAL.	
FT	SIGNAL	41	1151	M130 ANTIGEN, CYTOPLASMIC VARIANT 1.	
SO	SEQUENCE	1151 AA;	124820 NM;	A7ZED2F CRC32;	

	Query Match	Similarity	Score	DB #	Length
Best Local	51.0%	51.0%	380	49	1151
Matches	51	Conservative	Pred. No. 4,498-68	Mismatches 30	Indels 0
				Gaps 0	
Db	714	LRVNGGRCAGREITHEGSGWTCIDSDMLSDAHVVCQOLGCGEAINMTGSAHREGT	773		
447	VRLDDGNKKKEGRREVINGCWGTCIDGWTDDAAVICTGKGGARRRTAYPEEGK	506			


```

RESULT 13
ID 060997 PRELIMINARY: PRT: 2083 AA.
AC 060997:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CRP-DUCTIN PRECURSOR (CRP).
GN CRP OR CRP OR CRP-DUCTIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
SEQUENCE FROM N.A.
STRAIN-BALB/C; TISSUE=JEJUNUM;
MEDLINE: 96362470.
CHENG H., BJERKNES M., CHEN H.;
"CRP-ductin: a gene expressed in intestinal crypts and in pancreatic
and hepatic ducts.";
RL Anat. Rec. 244:327-343(1996).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN (CRP-ALPHA AND CRP-
BETA) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE
SEQUENCE SHOWN HERE IS THAT OF CRP-ALPHA.
CC -1- TISSUE SPECIFICITY: INTESTINE AND PANCREAS. IN THE COLON,
EXPRESSED IN CRYPT CELLS BUT NOT MATURE VILLOSUS CELLS. MOST HIGHLY
EXPRESSED IN MID-CRYPT. ALSO EXPRESSED IN EPITHELIUM LINING
HEPATIC AND PANCREATIC DUCTS.
CC EMBL: U37438; AAC52505.1; -.
DR HSP: P23932; ISEP.
DR MGD: MGI:106210; Crpd.
DR PFAM: PF00431; CUB; 5.
DR PFAM: PF00530; SRCR; 8.
DR PFAM: PF00100; zona_pellucida; 1.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SIGNAL: Transmembrane; Alternative splicing.
KW SIGNAL 1 28
FT CHAIN 29 2083 CRP-DUCTIN.
FT DOMAIN 29 2042 EXTRACELLULAR.
FT DOMAIN 986 1017 THR-RICH.
FT DOMAIN 1250 1259 PRO/THR-RICH.
FT DOMAIN 1610 1625 PRO/SER/THR-RICH.
FT TRANSMEM 2043 2065 POTENTIAL.
FT DOMAIN 2066 2083 CYTOPLASMIC.
FT VARSPLIC 2030 2083 LAVEDKPPASSQAVPTAIRIGVFLAMVLAFAFTIGRR
THIDGQPPSTKL -> PELSSQSGPSSH (IN
CRP-BETA).
SEQUENCE 2083 AA; 226734 MW; 945FC070 CRC32;
Query Match 48.7%; Score 369; DB 11; Length 2083;
Best Local Similarity 52.0%; Pred. No. 2,56e-65;
Matches 52; Conservative 16; Mismatches 32; Indels 0; Gaps 0;
Db 1508 LRLVNGSNCEGRVELLYGSGWTCVDDSGISDANVVCRLGCGSALSPGNAMFGGS 1567
:||||:|:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
Oy 447 VRLMDGKNEKRGVEVFINGWGITICDDGWTDKDAVICRLQGYGKPARAFMAYFGEG 506
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Db 1568 GLIVDDVSCSGEYLHNCCHRGMLVHNCRHSDEAGVIC 1607
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Oy 507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIG 546
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
RESULT 14
ID 0955X8 PRELIMINARY: PRT: 638 AA.
AC 0955X8:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE LYSYL OXIDASE-LIKE PROTEIN 2.
GN LOXL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
SEQUENCE FROM N.A.
RP TISSUE=SPLEEN, PLACENTA;
RC TISSUE=SPLEEN, PLACENTA;
RX MEDLINE: 99230328.
RA SAUV C.J., TRONECKER H., BOGIC L., BRYANT-GREENWOOD G.D., BOYD C.D.,
CSISZAR K.;
"the LOXL2 gene encodes a new lysyl oxidase-like protein and is
expressed at high levels in reproductive tissues.";
RL J. Biol. Chem. 274:12939-12944(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN, PLACENTA;
RA JOORDAN-LE SAUV C., CSISZAR K.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF117949; AAD34343.1; -.
DR PROSITE: PS00420; SPERACT RECEPTOR; 1.
SQ SEQUENCE 638 AA; 71101 MW; C232E6N4 CRC32;
Query Match 48.6%; Score 368; DB 4; Length 638;
Best Local Similarity 51.0%; Pred. No. 4,54e-65;
Matches 52; Conservative 21; Mismatches 26; Indels 3; Gaps 3;
Db 125 VALRGCAVIGEGRYVLKNGEWTCVDDKMDLVASVYCRELGF-GSAKEAVTGSRLGOG 183
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Oy 447 VRLMDGKNEKRGVEVFINGWGITICDDGWTDKDAVICRLQGYGKPAR-ARTMAYFGEG 505
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Db 184 IGPIHLNEIOCTGNKSIIDC-KFNAESOGCNHEEDAGRCN 224
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Oy 506 KGIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIG 547
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
RESULT 15
ID 09YAK0 PRELIMINARY: PRT: 774 AA.
AC 09YAK0:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE LYSYL OXIDASE-RELATED PROTEIN.
GN WS9-14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
SEQUENCE FROM N.A.
RA SATO H., PAPACONSTANTINOU J., SATO H., GOLDSTEIN S.;
"Regulation of a Novel Gene Encoding a Lysyl Oxidase-Related Protein
in Cellular Adhesion and Senescence.";
RL J. Biol. Chem. 0:0-0(1997).
DR EMBL: U89942; AAB49697.1; -.
DR PROSITE: PS00420; SPERACT RECEPTOR; 1.
SQ SEQUENCE 774 AA; 86724 MW; 1A5C47AA CRC32;
Query Match 48.6%; Score 368; DB 4; Length 774;
Best Local Similarity 51.0%; Pred. No. 4,54e-65;
Matches 52; Conservative 21; Mismatches 26; Indels 3; Gaps 3;
Db 326 VALRGCAVIGEGRYVLKNGEWTCVDDKMDLVASVYCRELGF-GSAKEAVTGSRLGOG 384
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Oy 447 VRLMDGKNEKRGVEVFINGWGITICDDGWTDKDAVICRLQGYGKPAR-ARTMAYFGEG 505
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Db 385 IGPIHLNEIOCTGNKSIIDC-KFNAESOGCNHEEDAGRCN 425
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|
Oy 506 KGIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAVIG 547
|||:||||:|:||||:| |||:||||:| |||:||||:| |||:||||:|

```

Search completed: Mon Mar 13 10:34:10 2000
Job time : 18 secs.

THIS PAGE BLANK (USPTO)

 RELEASE

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

ch-pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 13 10:33:26 2000; Maspar time 6.50 Seconds
 Tabular output not generated. 464,390 Million cell updates/sec

Title: >US-09-147-947-6
 Description: (447-547) from US09147947A.pep (6 of 6)
 Perfect Score: 757
 Sequence: 1 VALMDGEMKKEGVEVEFING.....KQDIGHNCRHSEADGAVICD 101

Scoring table: PAM 150
 Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 38.951; Variance 57.048; scale 0.683

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Hit No.	Score	Query Match	Length	ID	Description	Pred. No.
1	757	100.0	875	1	NETR_HUMAN NEUTROTYP SIN PRECURSOR	7.60e-177
2	724	95.6	761	1	NETR_MOUSE NEUTROTYP SIN PRECURSOR	1.50e-167
3	385	50.9	453	1	MSRE_BOVIN MACROPHAGE SCAVENGER R	5.00e-74
4	381	50.3	454	1	MSRE_RABIT MACROPHAGE SCAVENGER R	5.89e-73
5	368	48.6	458	1	MSRE_MOUSE MACROPHAGE SCAVENGER R	1.74e-69
6	362	47.8	1436	1	WC11_BOVIN ANTIGEN WC1.1	6.88e-68
7	361	47.7	451	1	MSRE_HUMAN MACROPHAGE SCAVENGER R	1.27e-67
8	351	46.4	532	1	SPER_STRU EGG PEPTIDE SPERACT RE	5.73e-65
9	323	42.7	468	1	CD6_HUMAN T-CELL DIFFERENTIATION	1.38e-57
10	157	20.7	1034	1	ENTK_PIG ENTEROPEPTIDASE PRECUR	2.57e-16
11	146	19.3	583	1	CFAI_HUMAN COMPLEMENT FACTOR 1 PR	7.47e-14
12	136	18.0	1069	1	ENTK_MOUSE ENTEROPEPTIDASE (EC 3.	1.15e-11
13	125	16.5	1035	1	ENTK_BOVIN ENTEROPEPTIDASE PRECUR	2.51e-09
14	120	15.9	1019	1	ENTK_HUMAN ENTEROPEPTIDASE PRECUR	2.73e-08
15	113	14.9	495	1	CD5_BOVIN T-CELL SURFACE GLYCOP	7.16e-07
16	106	14.0	492	1	TMS2_HUMAN TRANSMEMBRANE PROTEIN	1.71e-05
17	96	12.7	495	1	CD5_HUMAN T-CELL SURFACE GLYCOP	2.01e-03
18	95	12.5	494	1	CD5_MOUSE T-CELL SURFACE GLYCOP	1.06e-02
19	91	12.0	274	1	BPA1_STRU NON-HAEM BROMOPEROXIDA	1.59e-02
20	90	11.9	378	1	VR2_HP130 REGULATORY PROTEIN E2	7.84e-02
21	86	11.4	404	1	6P12_LYCS STILE DEVELOPMENT-SPEC	1.71e-01
22	84	11.1	1342	1	XDH_DROPS XANTHINE DEHYDROGENASE	1.71e-01
23	84	11.1	1344	1	XDH_DROSO XANTHINE DEHYDROGENASE	1.71e-01

24 83 11.0 650. 1 P72_HUMAN PROBABLE RNA-DEPENDENT 2.52e-01
 25 81 10.7 287. 1 NIFH_FRAP NITROGENASE IRON PROTE 5.38e-01
 26 81 10.7 305. 1 MIAA_AOUAE TRNA DELTA(2)-ISOPENTE 5.38e-01
 27 81 10.7 1335. 1 XDH_DROME XANTHINE DEHYDROGENASE 5.38e-01
 28 81 10.7 1758. 1 YR7_YEAST HYPOTHEICAL 197.5 KD 5.38e-01
 29 81 10.7 1758. 1 YW5_YEAST HYPOTHEICAL 197.5 KD 5.38e-01
 30 80 10.6 380. 1 KAP1_HUMAN CAMP-DEPENDENT PROTEIN 7.82e-01
 31 80 10.6 380. 1 KAP1_MOUSE CAMP-DEPENDENT PROTEIN 7.82e-01
 32 80 10.6 380. 1 KAP1_MOUSE CAMP-DEPENDENT PROTEIN 7.82e-01
 33 80 10.6 514. 1 ABF2_BACOV ALPHA-L-ARABINOFURANOS 7.82e-01
 34 79 10.4 507. 1 DAF_CAVPO COMPLEMENT DECAT-ACCEL 1.13e-00
 35 78 10.3 287. 1 NIFH_FRAP NITROGENASE IRON PROTE 1.64e-00
 36 78 10.3 514. 1 IMD2_MOUSE INOSINE-5'-MONOPHOSPA 1.64e-00
 37 77 10.2 453. 1 YKP9_KLULA HYDROXYMETHYLGUTARYL- 2.35e+00
 38 77 10.2 491. 1 HMCS_YEAST T-CELL SURFACE GLYCOP- 2.35e+00
 39 77 10.2 491. 1 CD5_RAT T-CELL SURFACE GLYCOP- 2.35e+00
 40 77 10.2 697. 1 YHFO_YEAST HYPOTHEICAL 79.0 KD P 2.35e+00
 41 77 10.2 1162. 1 VGL2_IVVA E2 GLYCOPROTEIN PRECUR 2.35e+00
 42 77 10.2 1363. 1 XDH_EMENT XANTHINE DEHYDROGENASE 2.35e+00
 43 76 10.0 192. 1 RPL1_BP186 REPRESSOR PROTEIN CI. 3.37e+00
 44 76 10.0 383. 1 METL_YEAST S-ADENOSYLMETHIONINE S 3.37e+00
 45 75 9.9 550. 1 MICO_STRCI MICOLISIN PRECURSOR (E 4.81e+00

RESULT 1 ALIGNMENTS
 ID NETR_HUMAN STANDARD: PRT; 875 AA.
 AC P67330;
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE NEUTROTYP SIN PRECURSOR (EC 3.4.21.-) (MOTOP SIN).
 GN PSS312.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SSEQUENCE FROM N.A.
 RC TISSUE-BRAIN: N.A.
 RX MEDLINE: 98201705.
 RA PROBA K., GSCHEWEND T.P., SONDEREGGER P.;
 RT "Cloning and sequencing of the cDNA encoding human neutrotypsin";
 RL Biochim. Biophys. Acta 1396:143-147(1998).
 CC -!- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
 ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
 LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 SRCR DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ001531; CA004816.1; -
 DR PROSITE: PS00134; TRYPSIN_HTS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS00420; SPERACT_RECEPTOR; 3.
 KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.
 FT CHAIN 1 20
 FT CHAIN 21 875
 FT DOMAIN 23 92
 FT DOMAIN 93 165
 FT DOMAIN 170 271
 FT DOMAIN 280 381
 FT DOMAIN 387 487
 FT SRCR 3.

SCR4
 Adv. Adv. 3.5 available

```

FT DOMAIN 500 601 SRCR 4
FT DOMAIN 619 875 SERINE PROTEASE.
FT DOMAIN 619 875 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 630 631 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM.
FT ACT_SITE 726 726 CHARGE RELAY SYSTEM.
FT ACT_SITE 825 825 CHARGE RELAY SYSTEM.
FT DISULFID 619 750 POTENTIAL.
FT CARBOHYD 26 26 POTENTIAL.
FT CARBOHYD 683 683 POTENTIAL.
SQ SEQUENCE 875 AA: 97011 MW: 67D5272B CRC32:

Query Match 100.0%: Score 757; DB 1; Length 875;
Best Local Similarity 100.0%: Pred. No. 7.67e-177;
Matches 101: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 500 VRLMDGKKEGKEVEFVINGOMGTICDDGWTDKDAAVICRQLGKGPARTMAFGECK 559
447 VRLMDGKKEGKEVEFVINGOMGTICDDGWTDKDAAVICRQLGKGPARTMAFGECK 506
560 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYICD 600
507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYICD 547

RESULT 2
ID NETR_MOUSE STANDARD: PRT: 761 AA.
AC 008762:
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE NEUTROTYPIN PRECURSOR (EC 3.4.21.-) (MOTOPIN) (BRAIN-SPECIFIC SERINE
PROTEASE 3) (BSSP-3).
RN PRSS12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97401523.
RA GSCHEIDT T.P., KRUEGER S.R., KOZLOV S.V., WOLFER D.P., SONDEREGGER P.;
RT "Neutrotypin, a novel multidomain serine protease expressed in the
nervous system.";
RL Mol. Cell. Neurosci. 9:207-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98008848.
RA YAMAMURA Y., YAMASHIRO K., TSURUOKA N., NAKAZATO H., TSUJIMURA A.,
RA YAMAGUCHI N.;
RT "Molecular cloning of a novel brain-specific serine protease with a
kringle-like structure and three scavenger receptor cysteine-rich
motifs.";
RL Biochem. Biophys. Res. Commun. 239:386-392(1997).
CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTEOLYTIC
ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH
LEARNING AND MEMORY OPERATIONS.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN CEREBRAL CORTEX, HIPPOCAMPUS
AND AMYGDALA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Y13192; CA473646.1; -
DR EMBL: D89871; BA423986.1; -
DR MGD: MGI:1100861; PRSS12.
DR PFAM: PF00530; SRCR_3.
DR PFAM: PF00089; trypsin; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00420; SPERACT-RECEPTOR; 3.
KW Hydrolase; Serine protease; Glycoprotein; Kringle; Repeat; signal.
FT SIGNAL 1 21
FT CHAIN 22 761 NEUTROTYPIN.
FT DOMAIN 85 157 KRINGLE.
FT DOMAIN 166 267 SRCR 1.
FT DOMAIN 273 373 SRCR 2.
FT DOMAIN 386 487 SRCR 3.
FT DOMAIN 505 761 SERINE PROTEASE.
FT DOMAIN 505 516 ZMOGEN ACTIVATION REGION.
FT ACT_SITE 516 517 REACTIVE BOND (POTENTIAL).
FT ACT_SITE 562 562 CHARGE RELAY SYSTEM.
FT ACT_SITE 612 612 CHARGE RELAY SYSTEM.
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM.
FT DISULFID 505 636 POTENTIAL.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 521 521 POTENTIAL.
FT CARBOHYD 569 569 POTENTIAL.
SQ SEQUENCE 761 AA: 84118 MW: 3F3C4F35 CRC32:

Query Match 95.6%: Score 724; DB 1; Length 761;
Best Local Similarity 92.1%: Pred. No. 1.50e-167;
Matches 93: Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 386 IRLVDGKKEGKEVEFVINGOMGTICDDGWTDKDAAVICRQLGKGPARTMAFGECK 445
447 VRLMDGKKEGKEVEFVINGOMGTICDDGWTDKDAAVICRQLGKGPARTMAFGECK 506
446 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYICD 486
507 GPIHVDNVKCTGNERSLADCIKODIGRHNCRHSEDAGVYICD 547

RESULT 3
ID MSRE_BOVIN STANDARD: PRT: 453 AA.
AC P21758;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
IDL RECEPTOR I AND II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-LUNG;
RX MEDLINE: 90136965.
RA KODAMA T., FREEMAN M., ROHRER L., ZABRECKY J., MATSUDAIRA P.,
RA KRIEGER M.;
RT "Type I macrophage scavenger receptor contains alpha-helical and
collagen-like coiled coils.";
RL Nature 343:531-535(1990).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE-LUNG;
RX MEDLINE: 90136973.
RA ROHRER L., FREEMAN M., KODAMA T., PENMAN M., KRIEGER M.;
RT "Coiled-coil fibrous domains mediate ligand binding by macrophage
scavenger receptor type II.";
RL Nature 343:570-572(1990).
CC -1- FUNCTION: MEMBRAN GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC
DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING

```

CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT. HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X51689; CAA35987.1; -
 CC EMBL: X54183; CAA38108.1; -
 CC PIR: S08276; S08276.
 CC PIR: S08278; S08278.
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR PFAM: PF01391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; 1.
 CC KM Heptad repeat pattern; Receptor; Alternative splicing.
 CC FT DOMAIN 1 50
 CC FT TRANSMEM 1 76
 CC FT DOMAIN 77 453
 CC FT DOMAIN 77 108
 CC FT DOMAIN 109 271
 CC FT DOMAIN 272 343
 CC FT DOMAIN 344 453
 CC FT CARBOHYD 82 82
 CC FT CARBOHYD 101 101
 CC FT CARBOHYD 142 142
 CC FT CARBOHYD 183 183
 CC FT CARBOHYD 220 220
 CC FT CARBOHYD 248 248
 CC FT CARBOHYD 266 266
 CC FT VARSPLIC 348 349
 CC FT VARSPLIC 350 453
 CC SQ SEQUENCE 453 AA; 50056 MW; 2CDE1EE1 CRC32;
 CC -----
 CC Query Match 50.98; Score 385; DB 1; Length 453;
 CC Best Local Similarity 50.08; Pred. No. 5,00e-74;
 CC Matches 50; Conservative 17; Mismatches 33; Indels 0; Gaps 0;
 CC -----
 CC 352 VRLVGGSGPHEGRVEIFHEGOMGTVCDDRMELGLVCRSLGYKGYSHKRAYGKGT 411
 CC 447 VRLMDGNKKEGREVFINGOMGTICDDGWTDKDAVYICQDLYKGPAPARATMAIFGEGR 506
 CC DB 412 GPIWNEVPCFGESEIECKIRQWGVYACSHGSDAGVTC 451
 CC QY 507 GPIHVDNVKCTGNERSIADCIKQDIGHNCRHSESDAGVIC 546
 CC -----
 CC RESULT 4
 CC ID MSRE_RABIT STANDARD; PRT; 454 AA.
 CC AC 00585;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC LDL RECEPTOR I AND II).
 CC GN MSRL.
 CC OS Oryctolagus cuniculus (Rabbit).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC RN Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 93016877.
 CC RA BICKEL P.E., FREEMAN M.W.;
 CC RT "Rabbit aortic smooth muscle cells express inducible macrophage
 CC scavenger receptor messenger RNA that is absent from endothelial

FT cells".
 RL J. Clin. Invest. 90:1450-1457(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93131972.
 RA DOI T., MADA Y., KODAMA T., HIGASHI K.I., KURIHARA Y.,
 RA MIYAZAKI T., NAKAMURA H., UESUGI S., IMANISHI T., KAMABE Y.,
 RA ITAKURA H., YAZAKI Y., MATSUMOTO A.;
 RT "Charged collagen structure mediates the recognition of negatively
 RT charged macromolecules by macrophage scavenger receptors".
 RL J. Biol. Chem. 268:2126-2133(1993).
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEIN. IMPLICATED IN THE PATHOLOGIC
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -1- SUBUNIT. HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: THE TWO FORMS OF MACROPHAGE SCAVENGER
 CC RECEPTOR (TYPES I AND II) ARE PRODUCED BY ALTERNATIVE SPLICING
 CC OF THE SAME GENE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L11693; AAA31402.1; -
 CC EMBL: L11692; AAA31403.1; -
 CC EMBL: D13381; BAA02649.1; -
 CC DR PROSITE: PS00420; SPERACT_RECEPTOR; 1.
 CC DR PFAM: PF00530; SRCR; 1.
 CC DR PFAM: PF01391; Collagen; 1.
 CC KM Transmembrane; Glycoprotein; 1.
 CC KM Heptad repeat pattern; Receptor; Alternative splicing.
 CC FT DOMAIN 1 50
 CC FT TRANSMEM 1 73
 CC FT DOMAIN 74 454
 CC FT DOMAIN 74 109
 CC FT DOMAIN 110 272
 CC FT DOMAIN 273 344
 CC FT DOMAIN 345 454
 CC FT CARBOHYD 82 82
 CC FT CARBOHYD 102 102
 CC FT CARBOHYD 143 143
 CC FT CARBOHYD 184 184
 CC FT CARBOHYD 221 221
 CC FT CARBOHYD 249 249
 CC FT CARBOHYD 267 267
 CC FT VARSPLIC 348 354
 CC FT VARSPLIC 355 454
 CC FT CONFLICT 106 106
 CC SQ SEQUENCE 454 AA; 49745 MW; 5D780348 CRC32;
 CC -----
 CC Query Match 50.38; Score 381; DB 1; Length 454;
 CC Best Local Similarity 49.08; Pred. No. 5,89e-73;
 CC Matches 49; Conservative 20; Mismatches 31; Indels 0; Gaps 0;
 CC -----
 CC 353 VRLVGGSGPHEGRVEIFHEGOMGTVCDDRMELRAGOVYCRSVHKKAYGKGT 412
 CC 447 VRLMDGNKKEGREVFINGOMGTICDDGWTDKDAVYICQDLYKGPAPARATMAIFGEGR 506
 CC DB 413 GPIWNEVPCFGESEIECKIRQWGVYACSHGSDAGVTC 452
 CC QY 507 GPIHVDNVKCTGNERSIADCIKQDIGHNCRHSESDAGVIC 546
 CC -----
 CC RESULT 5
 CC ID MSRE_MOUSE STANDARD; PRT; 458 AA.
 CC AC 00585;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE MACROPHAGE SCAVENGER RECEPTOR TYPES I AND II (MACROPHAGE ACETYLATED
 CC LDL RECEPTOR I AND II).
 CC GN MSRL.
 CC OS Oryctolagus cuniculus (Rabbit).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC RN Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 93016877.
 CC RA BICKEL P.E., FREEMAN M.W.;
 CC RT "Rabbit aortic smooth muscle cells express inducible macrophage
 CC scavenger receptor messenger RNA that is absent from endothelial

	FT	CARBOHYD	221	221	POTENTIAL.
	FT	CARBOHYD	249	249	POTENTIAL.
	FT	CARBOHYD	267	267	POTENTIAL.
	FT	VARSPLIC	345	358	TTPFKATLVGSGP -> REVOLTDHRAGPS (IN ISOFORM II).
	FT	VARSPLIC	359	451	MISSING (IN ISOFORM II).
SQ	SEQUENCE	451 AA;	49762 MW;	AB9FEECT CRC32;	
	Query Match				
	Best Local Similarity		47.7%;	Score 361;	DB 1:
Matches	49;	Conservative	17;	Mismatches 34;	Indels 0;
				Gaps 0;	
Df	350	VRVLGGSPHEGRVPEILHSQGWTICDDRWKRVRYGVYCNLSLGPVQAVAHKAHFGOGT	409		
Oy	447	VRLDMDGENKKRGRRVEFLNNGWGTICDDGWKDPAAYVICGYKGAPARTMAVFEGEK	506		
Df	410	GPIWMENFCEGFRESSIECKIRBMGRACSHSDAGVC	449		
Oy	507	GPIHDNVKCTGNERSIADCIKIDIGHRNCNHSDEDAGVIC	546		
RESULT	8	STANDARD;	PRT;	532 AA.	
ID	SPER_STRPU	STDAND;	PRT;	532 AA.	
AC	p16264;				
DT	01-AUG-1990	(Rel. 15,	Created)		
DT	01-AUG-1990	(Rel. 15,	Last sequence update)		
DT	15-JUL-1999	(Rel. 38,	Last annotation update)		
DE	EGG PEPTIDE SPERRACT RECEPTOR PRECURSOR.				
OS	Strongylocentrotus purpuratus (Purple sea urchin).				
OC	Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidae; Euechinoidea; Echnidae; Echinoida; Strongylocentrotidiae; Strongylocentrotus.				
CC	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 477-489.				
RX	MEDLINE, 89184581.				
RA	DARGOTT U.J., JORDAN J.E., BELLET R.A., GARBES D.L.;				
RT	"Cloning of the mRNA for the protein that crosslinks to the egg peptide speract.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2128-2132(1989).				
CC	-1- FUNCTION: RECEPTOR FOR THE EGG PEPTIDE SPERRACT.				
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; J04518; AAA30078.1;				
DR	PIR: A32751.				
DR	PROSITE, PS00420: SPERRACT_RECEPTOR; 4.				
DR	PFAM: PF00530; SSCR; 4.				
KW	Receptor; Transmembrane; glycoprotein; signal; Repeat.				
FT	SIGNAL	1	30		
FT	CHAIN	31	532	EGG PEPTIDE SPERRACT RECEPTOR.	
FT	DOMAIN	31	491	EXTRACELLULAR (POTENTIAL).	
FT	TRANSERM	492	520	POTENTIAL.	
FT	DOMAIN	521	532	CYTOSOLSMTC (POTENTIAL).	
FT	REPEAT	40	147	4 X APPROXIMATE TANDEM REPEATS.	
FT	REPEAT	148	260	1.	
FT	REPEAT	261	375	2.	
FT	REPEAT	376	488	3.	
FT	CARBOHYD	78	78	4.	
FT	CARBOHYD	115	115	POTENTIAL.	
FT	CARBOHYD	459	459	POTENTIAL.	
SQ	SEQUENCE	532 AA;	57820 MW;	D27174AE CRC32;	
	Query Match				
	Best Local Similarity		46.4%;	Score 351;	DB 1:
Matches	53;	Conservative	16;	Mismatches 30;	Indels 4;
				Gaps 4;	

```

Db 264 IILMDSGGHEERVELIWMHDAGTICDDCDMDADANAVCRQAGYRQAVASFGKGEFGF 323
QY 447 VALMOGKKEKBERVEVFIINGCMGTICDDCDMDTCKDAVIRQLGKYPAPART-MAY-FSE 504
Db 324 TWAPHTSFVCTGVEDRLDICTLRD-GWTHSCYHEDASVVC 365
QY 505 GKGPIHVDNWKCTGNERSLADCIKODIGR-HNCRHSEDAGVIC 546

RESULT 9
ID CD6_HUMAN STANDARD: PRT: 468 AA.
AC P30203:
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE T-CELL DIFFERENTIATION ANTIGEN CD6 PRECURSOR (I12) (TP120).
GN CD6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homidae; Homo.
[1]
SEQUENCE FROM N.A.
MEDLINE: 92013796.
RX ARUFFO A., MELNICK M.B., LINSLEY P.S., SEED B.;
RA "The lymphocyte glycoprotein CD6 contains a repeated domain structure
RT characteristic of a new family of cell surface and secreted
RT proteins".
RL J. Exp. Med. 174:949-952(1991)
CC - FUNCTION: INVOLVED IN CELL ADHESION. BINDS TO CD166.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC - SIMILARITY: CONTAINS 3 SRCR DOMAINS.
CC - DATABASE: NAME=PROV; NOTE=CD guide cd6 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd6.htm"
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60992; CAA43306.1; .
CC DR PIR: S26741; S26741.
CC DR MIM: 186720; .
CC DR PROSITE: PS00420; SPERACT_RECEPTOR: 1.
CC DR PFAM: PF00530; SRCR: 3.
CC KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal.
CC FT SIGNAL 1 ?
CC FT CHAIN ?
CC FT DOMAIN ? 468 T-CELL DIFFERENTIATION ANTIGEN CD6.
CC FT TRANSMEM 403 423 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 424 468 CYTOPLASMIC (POTENTIAL).
CC FT CARBOHYD 28 28 POTENTIAL.
CC FT CARBOHYD 49 49 POTENTIAL.
CC FT CARBOHYD 112 112 POTENTIAL.
CC FT CARBOHYD 118 118 POTENTIAL.
CC FT CARBOHYD 229 229 POTENTIAL.
CC FT CARBOHYD 339 339 POTENTIAL.
CC FT CARBOHYD 345 345 POTENTIAL.
CC FT CARBOHYD 368 368 POTENTIAL.
CC FT SEQUENCE 468 AA; 49828 MW; 042D55AC8 CRC32;

Query Match 42.7% Score 323; DB 1; Length 468;
Best Local Similarity 48.0%; Pred. No. 1,38e-57;
Matches 46; Conservative 16; Mismatches 35; Indels 1; Gaps 1;

Db 161 LRLVGGGACAGARVEMLEHGEWGVCDTWDLEDANAVCRQLGCGMAVALPGLHFTPTGR 220
QY 447 VRLMDGKKEKBERVEVFIINGCMGTICDDCDMDTCKDAVIRQLGKYPAPARTIMAFGEGR 506
Db 221 GPIHDDQVNCGAEAYLMDCPGLP-GQHYCGKEDAGVVC 259

```



```

FT DOMAIN 197 238 LDI-RECEPTOR CLASS A 1.
FT DOMAIN 240 349 CUB.
FT DOMAIN 357 519 MAM.
FT DOMAIN 539 649 CUB.
FT DOMAIN 656 694 LDI-RECEPTOR CLASS A 2.
FT DOMAIN 693 786 SRCR.
FT ACT_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 236 BY SIMILARITY.
FT DISULFID 658 670 BY SIMILARITY.
FT DISULFID 665 683 BY SIMILARITY.
FT DISULFID 677 692 BY SIMILARITY.
FT DISULFID 787 911 INTERCHAIN (BY SIMILARITY).
FT DISULFID 825 841 BY SIMILARITY.
FT DISULFID 925 992 BY SIMILARITY.
FT DISULFID 956 971 BY SIMILARITY.
FT DISULFID 982 1010 BY SIMILARITY.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 170 170 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT CARBOHYD 283 283 POTENTIAL.
FT CARBOHYD 343 343 POTENTIAL.
FT CARBOHYD 350 350 POTENTIAL.
FT CARBOHYD 403 403 POTENTIAL.
FT CARBOHYD 455 455 POTENTIAL.
FT CARBOHYD 485 485 POTENTIAL.
FT CARBOHYD 518 518 POTENTIAL.
FT CARBOHYD 549 549 POTENTIAL.
FT CARBOHYD 645 645 POTENTIAL.
FT CARBOHYD 697 697 POTENTIAL.
FT CARBOHYD 701 701 POTENTIAL.
FT CARBOHYD 721 721 POTENTIAL.
FT CARBOHYD 740 740 POTENTIAL.
FT CARBOHYD 761 761 POTENTIAL.
FT CARBOHYD 804 804 POTENTIAL.
FT CARBOHYD 863 863 POTENTIAL.
FT CARBOHYD 902 902 POTENTIAL.
FT CARBOHYD 964 964 POTENTIAL.
SQ SEQUENCE 1034 AA; 114776 MW; 24386471 CRC32;

ery Match 20.7%; Score 157; DB 1; Length 1034;
st.Local Similarity 31.7%; Pred.No.2.57e-16;
lches 20; Conservative 18; Mismatches 23; Indels 2; Gaps 2;

```

```

RA CATERALL C.F., LYONS A., SIM R.M., DAY A.J., HARRIS T.J.R.;
RT "Characterization of primary amino acid sequence of human complement
RT control protein factor I from an analysis of cDNA clones.";
RN Biochem. J. 242:849-856(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87280021.
RA GOLDBERGER G., BRUNS G.A.P., RITS M., EDGE M.D., KWATROWSKI D.J.;
RT "Human complement factor I: analysis of cDNA-derived primary
RT structure and assignment of its gene to chromosome 4.";
RT J. Biol. Chem. 262:10065-10071(1987).
CC -1- FUNCTION: FACTOR I IS RESPONSIBLE FOR CLEAVING THE ALPHA-CHAINS
CC OF C4B AND C3B IN THE PRESENCE OF THE COFACTORS C4-BINDING
CC PROTEIN AND FACTOR H RESPECTIVELY.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT AND HEAVY CHAINS LINKED BY
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: PLASMA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
CC EMBL: Y00318; CAA68416.1; ALT_INIT.
CC EMBL: J02770; AAA52455.1; -.
CC PIR: A29154; A29154.
CC HSSP: P00763; IDPO.
CC SWISS-2DPAGE: P05156; HUMAN.
CC MIM: 217030; -.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01209; LDLRA_1; 1.
CC PROSITE: PS0068; LDLRA_2; 2.
CC PFM: PFO0057; ldl_recept_a; 2.
CC PFM: PFO0089; trypsin; 1.
CC PFM: PFO0530; SRCR; 1.
CC Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 583 COMPLEMENT FACTOR I, HEAVY CHAIN.
FT CHAIN 19 335 COMPLEMENT FACTOR I LIGHT CHAIN.
FT CHAIN 340 583 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 220 258 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 257 295 BY SIMILARITY.
FT DISULFID 229 247 BY SIMILARITY.
FT DISULFID 241 259 BY SIMILARITY.
FT DISULFID 259 271 BY SIMILARITY.
FT DISULFID 266 284 BY SIMILARITY.
FT DISULFID 278 293 BY SIMILARITY.
FT DISULFID 365 381 BY SIMILARITY.
FT DISULFID 467 531 BY SIMILARITY.
FT DISULFID 495 510 BY SIMILARITY.
FT DISULFID 521 550 BY SIMILARITY.
FT ACT_SITE 380 380 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 429 429 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 525 525 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 70 70 POTENTIAL.
FT CARBOHYD 103 103 POTENTIAL.
FT CARBOHYD 177 177 POTENTIAL.
FT CARBOHYD 464 464 POTENTIAL.
FT CARBOHYD 494 494 POTENTIAL.
FT CARBOHYD 536 536 POTENTIAL.
FT CONFLICT 558 558 V -> F (IN REF. 2).
SQ SEQUENCE 583 AA; 65720 MW; D673E9F0 CRC32;

Query Match 19.3%; Score 146; DB 1; Length 583;

```


RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 RN [2]
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 9404312.
 RA LAVALLIE E.R., REHEMUTULA A., RACIE L.A., DIBLASIO E.A.,
 RA FERENZ C., GRANT K.L., LIGHT A., MCCOY J.M.;
 RT "Cloning and functional expression of a cDNA encoding the catalytic
 RT subunit of bovine enterokinase.";
 RL J. Biol. Chem. 268:23311-23317(1993).
 RN [3]
 RP SEQUENCE OF 801-827.
 RX TISSUE-INTESTINE;
 RA MEDLINE: 92189715.
 RA LIGHT A., JANSKA H.;
 RT "The amino-terminal sequence of the catalytic subunit of bovine
 RT enterokinase.";
 RL J. Protein Chem. 10:475-480(1991).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U09839; AAB40026.1; -
 CC EMBL: L19663; AAL16035.1; -
 CC PIR: A61436; A61436.
 CC HSSP: P00763; IDPO.
 CC PROSITE: PS00134; TRYPSIN_HTS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS50060; MAM_2; 1.
 CC PROSITE: PS01209; LDLRA_1; 2.
 CC PROSITE: PS50068; LDLRA_2; 2.
 CC PFAM: PF00057; ldl_recept_a; 2.
 CC PFAM: PF00089; trypsin; 1.
 CC PFAM: PF00431; CUB; 2.
 CC PFAM: PF00530; SRCR; 1.
 CC PFAM: PF00629; MAM; 1.
 CC PFAM: PF01390; SEA; 1.
 CC Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 CC Setine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 CC CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 CC TRANSMEM 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
 CC DOMAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 CC DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 CC DOMAIN 240 350 CUB.
 CC DOMAIN 358 520 MAM.
 CC DOMAIN 340 650 CUB.

FT DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
 FT ACT_SITE 694 787 SRCR.
 FT ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT LIPID 2 2 MYRISTATE (POTENTIAL).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 219 236 BY SIMILARITY.
 FT DISULFID 659 671 BY SIMILARITY.
 FT DISULFID 666 684 BY SIMILARITY.
 FT DISULFID 678 693 BY SIMILARITY.
 FT DISULFID 768 912 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 826 842 BY SIMILARITY.
 FT DISULFID 926 993 BY SIMILARITY.
 FT DISULFID 957 972 BY SIMILARITY.
 FT DISULFID 983 1011 BY SIMILARITY.
 FT CARBOHYD 116 116 POTENTIAL.
 FT CARBOHYD 147 147 POTENTIAL.
 FT CARBOHYD 170 170 POTENTIAL.
 FT CARBOHYD 194 194 POTENTIAL.
 FT CARBOHYD 233 233 POTENTIAL.
 FT CARBOHYD 263 263 POTENTIAL.
 FT CARBOHYD 264 264 POTENTIAL.
 FT CARBOHYD 404 404 POTENTIAL.
 FT CARBOHYD 456 456 POTENTIAL.
 FT CARBOHYD 486 486 POTENTIAL.
 FT CARBOHYD 519 519 POTENTIAL.
 FT CARBOHYD 550 550 POTENTIAL.
 FT CARBOHYD 646 646 POTENTIAL.
 FT CARBOHYD 698 698 POTENTIAL.
 FT CARBOHYD 722 722 POTENTIAL.
 FT CARBOHYD 741 741 POTENTIAL.
 FT CARBOHYD 762 762 POTENTIAL.
 FT CARBOHYD 864 864 POTENTIAL.
 FT CARBOHYD 903 903 POTENTIAL.
 FT CARBOHYD 965 965 POTENTIAL.
 FT VARSPLIC 166 192 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 808 808 R -> Y (IN REF. 3).
 SQ SEQUENCE 1035 AA; 114867 MM; 5173034A CMC32;
 Query Match 16.5%; Score 125; DB 1; Length 1035;
 Best Local Similarity 30.6%; Pred. No. 2,51e-09;
 Matches 19; Conservative 13; Mismatches 28; Indels 2; Gaps 2;
 Db 694 VALFNQTTSSGLNFERIOSIHVACAEWNTQISDVQVLGL-GTGN-SVPIFSTGG 751
 QY 447 VRLMGENKREGRVEYFINGQMTICDDGMDKDAVTCROLDYKGPARRATMAYFGEK 506
 Db 752 GP 753
 QY 507 GP 508
 RESULT 14
 ID ENTK_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33; Created)
 DT 01-FEB-1996 (Rel. 33; Last sequence update)
 DT 15-JUL-1998 (Rel. 36; Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A..
 RC TISSUE=DUDENUM;
 RX MEDLINE: 95234679.
 RA KITAMOTO Y., VEILE R.A., DONIS-KELLER H., SADLER J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 RT the proteolytic activator of trypsinogen.";
 RL Biochemistry 34:4562-4568(1995).

FT	DISULFID	650	668	BY SIMILARITY.
FT	DISULFID	662	677	BY SIMILARITY.
FT	DISULFID	772	896	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	810	826	BY SIMILARITY.
FT	DISULFID	910	977	BY SIMILARITY.
FT	DISULFID	941	956	BY SIMILARITY.
FT	DISULFID	967	995	BY SIMILARITY.
FT	CARBOHYD	116	116	POTENTIAL.
FT	CARBOHYD	147	147	POTENTIAL.
FT	CARBOHYD	179	179	POTENTIAL.
FT	CARBOHYD	328	328	POTENTIAL.
FT	CARBOHYD	335	335	POTENTIAL.
FT	CARBOHYD	388	388	POTENTIAL.
FT	CARBOHYD	440	440	POTENTIAL.
FT	CARBOHYD	470	470	POTENTIAL.
FT	CARBOHYD	503	503	POTENTIAL.
FT	CARBOHYD	534	534	POTENTIAL.
FT	CARBOHYD	630	630	POTENTIAL.
FT	CARBOHYD	682	682	POTENTIAL.
FT	CARBOHYD	706	706	POTENTIAL.
FT	CARBOHYD	725	725	POTENTIAL.
FT	CARBOHYD	848	848	POTENTIAL.
FT	CARBOHYD	887	887	POTENTIAL.
FT	CARBOHYD	909	909	POTENTIAL.
FT	CARBOHYD	949	949	POTENTIAL.
SQ	SEQUENCE	1019 AA;	112923 MW; OE64IC53 CRC32;	
 Query Match 15.9%; Score 120; DB 1; Length 1019; Best Local Similarity 25.5%; Pred. No. 2,73e-08; Matches 24; Conservative 23; Mismatches 43; Indels 4; Gaps 4;				
Dd	678 VREFNGTNNGLVRFERISQIWHYCAENWTTQISNDVCQLLGI-GSGNSKPIFFTDG- 735			
Oy	447 VRLMDGENKKKEVEVFINGWGITCIDDDGWTDKDAIVAPOLGYKGPARATMYAREGGK 506			
Dd	736 GPFYKLNTARDGHLLITPSOCLQDSLRLOCNH 769			
Oy	507 GPIHDNVKCTGNR-SLAD-CIKODIGHNCRH 538			
 RESULT 15 AC CD5 BOVIN STANDARD; PRT; 495 AA. AD P19238; DT 01-NOV-1990 (Rel. 16, Created) DT 01-NOV-1990 (Rel. 16, Last sequence update) DT 01-NOV-1997 (Rel. 35, Last annotation update) DE T-CELL SURFACE GLYCOPROTEIN CD5 PRECURSOR (LYMPHOCTE GLYCOPROTEIN DE TI/LCB-1) (LYMPHOCTE ANTIGEN CD5). GN CD5. OS Bos taurus (Bovine). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; OC Euthelia; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; OC Bovinae; Bos. CC [1] RP SEQUENCE FROM N.A. RC TISSUE=THYMUS; RX MEDLINE: 90384847. RA YU Q., REICHERT M., BROUSSEAU T., CLEUTER Y., BURNY A., KETTMANN R.; RT "Sequence of bovine CD5."; RL Nucleic Acids Res. 18:5296-5296(1990). CC -!- FUNCTION: MAY ACT AS A RECEPTOR IN REGULATING T-CELL PROLIFERATION. CD5 INTERACTS WITH CD72/TYB-2. CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CC -!- SIMILARITY: CONTAINS 2 SRP DOMAINS. CC ----- CC This SWISS-Prot entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ CC or send an email to license@isb-sib.ch). CC -----				

•
•
•

495 AA; 54333 MW; F8ED2AE0 CRC32;

Conservative 8; Mismatches 12; Indels 4; Gaps 4;

318 CREVEKAR IANKMANJEDDCEWVGVGKNSVCEG3

ENKKEGVEVI-NGQ-WGTICDDGWTDKDAA-V-ICRQ 487

CS.

THIS PAGE BLANK (USPTO)